Association of Human and Machine Segmentation Using Blurred Images

Maria Lena Demetriou

Bachelor of Science in Computer Science with Honours
The University of Bath
May 2010

Association of Human and Machine Segmentation Using Blurred Images

Association of Human and Machine Segmentation Using Blurred Images

Submitted by: Maria Lena Demetriou

COPYRIGHT

Declaration

This dissertation is submitted to the University of Bath in accordance with the requirements of the degree of Bachelor of Science in the Department of Computer Science. No portion of the work in this dissertation has been submitted in support of an application for any other degree or qualification of this or any other university or institution of learning. Except where specifically acknowledged, it is the work of the author.

Signed:

Abstract

Based on Malik et al. research in 2004 (Berkeley) on the association of human and machine segmentation for natural images, project extends research to blurred images. Its main aim is to observe the effects of the use of blurred images on both human and machine segmentation. Project sets out to test two hypotheses. The first one relates to association in machine and human segmentation performance as the images' blurring level increases, while the second one analyses images' human cognition. Image segmentation on blurred images has not been analysed ever before; therefore this project not only shall provide ground truth of human segmentations on blurred images, but also initialise a new area of image segmentation research.

Contents

CONTENTS .... **I**

LIST OF FIGURES .... **III**

LIST OF TABLES .... **IV**

ACKNOWLEDGEMENTS .... **V**

INTRODUCTION **1**

1.1 AIM 2
1.2 OBJECTIVES 3
Hypothesis I *3*
Hypothesis 2: *3*
1.3 RESEARCH 4

LITERATURE SURVEY **5**

2.1 IMAGE SEGMENTATION 6
2.1.1 Definition *6*
2.1.2 Human – Machine Segmentation Comparison *6*
2.2 IMAGE FEATURES OF BOUNDARY DETECTION 7
2.2.1 Gradient-Based Features *7*
2.2.2 Oriented Energy *7*
2.2.3 Colour Gradient *8*
2.2.4 Brightness Gradient *8*
2.2.5 Texture Gradient *9*
2.2.6 Localisation *10*
2.3 EVALUATION METHODOLOGY 12
2.3.1 Precision-Recall Curve *12*
2.4 CUE OPTIMISATION 14
2.5 CUE COMBINATION 15
2.5.1 Classifiers *16*
2.6 PROGRAMMING BACKGROUND 17
2.6.1 Blurring Images Mechanism *17*
2.6.2 Machine Segmentor *17*
2.6.3 Human Segmentation Support Interface *17*
2.6.4 Results Presentation Interface *18*
2.6.5 Graphical User Interfaces using Matlab *18*
2.6.6 Graphical User Interfaces for Research *18*
2.7 EXPECTATIONS 19
2.8 SUMMARY 19

DESIGN OF EXPERIMENTS **21**

EXPERIMENTAL RESULTS 27

ANALYSIS OF RESULTS 36

5.1. HUMAN AND MACHINE PERFORMANCE 36
5.2. HUMAN COGNITION OF IMAGES 38

CONCLUSIONS 40

- *Human Segmentation Experiments Approach II 43*
- *Human Segmentation Experiments Final Approach 43*

BIBLIOGRAPHY 51

RESEARCH ETHICS 54

CODE 57

BLURRING CODE 57
MACHINE SEGMENTATION AND EVALUATION 58

List of Figures

FIGURE 1. BRAIN SEGMENTATION MEDICAL IMAGE [3] 2
FIGURE 2. COMPUTING TEXTONS. (A) FILTER BANK: THE 13-ELEMENT FILTER BANK USED FOR COMPUTING TEXTONS. (B) UNIVERSAL TEXTONS (C) IMAGE AND (D) TEXTON MAP [6] 9
FIGURE 3. (A) RAW OE, (B) RAW BG, (C) RAW CG, (D) RAW TG, (E) LOCALIZED OE, (F) LOCALIZED BG, (G) LOCALIZED CG, (H) LOCALIZED TG 11
FIGURE 4. TWO DECADES OF BOUNDARY DETECTION 13
FIGURE 5. BRIGHTNESS GRADIENT, (B) COLOUR GRADIENT, KERNEL BANDWIDTH FOR BG AND CG KERNEL DENSITY ESTIMATES 14
FIGURE 6. (A) ORIGINAL IMAGE, (B) BLURRED IMAGE WITH Σ=8, (C) BLURRED IMAGE WITH Σ=16, (D) BLURRED IMAGE WITH Σ=32 22
FIGURE 7. (A) ORIGINAL IMAGE, (B) BLURRED IMAGE WITH MATRIX SIZE 5, (C) BLURRED IMAGE WITH MATRIX SIZE 10, (D) BLURRED IMAGE WITH MATRIX SIZE 15 23
FIGURE 8. MACHINE SEGMENTATION ON AN IMAGE OF ZERO BLURRINESS: 25
FIGURE 9. EXAMPLE HUMAN SEGMENTATION FOR ALL FOUR LEVELS OF BLURRINESS (FROM ZERO TO MAXIMUM) IN COLOUR AND GREYSCALE FORM 28
FIGURE 10. MACHINE SEGMENTATIONS ON A COLOURED ORIGINAL IMAGE (LEFT) AND A COLOURED MAXIMUM BLURRED IMAGE (RIGHT) 29
FIGURE 11. COMBINED PRECISION RECALL GRAPHS 30
FIGURE 12. MATCHES BETWEEN MACHINE AND HUMAN SEGMENTATIONS FOR ALL FOUR LEVELS OF BLURRINESS 31
FIGURE 13. PR GRAPHS FOR ORIGINAL AND BLURRED FORMATS. 32
FIGURE 14. ORIGINAL IMAGE AND MINIMUM BLURRED IMAGE. 32
FIGURE 15. MEDIUM BLURINESS IMAGE AND MAXIMUM BLURINESS IMAGE. 33
FIGURE 16. HUMAN SEGMENTATIONS ON TWO IMAGES. 34

List of Tables

TABLE 1 RESULTS GATHERED FROM QUESTIONNAIRES.. 35

Acknowledgements

I would like to thank my supervisor, Dr Peter Hall for the support and guidance he offered over the past year. I would also like to thank David Martin, Charless Fowlkes and Juan Cardelino for the interest they showed and for their advices and suggestions.

I would also like to say a huge thank you to all of the experiments' participants, for their commitment and time they dedicated to this project.

Chapter 1

Introduction

This project was based on a computer vision area, that of image segmentation. Its main purpose was to examine the difference between machine and human segmentation performance. This notion is not new in the area of computer vision, but on the contrary this evaluation of difference has already been widely researched. Many machine segmentors have been developed over the years, but until today, none of them achieved performance similar to those of humans. This even stands for the best segmentor, which despite the fact it was implemented in 2004 by research team of University of Berkeley in California, is still majorly used. While this segmentor performs better than any other machine segmentor, it has significantly lower performance in comparison to human segmentation. Efficient machine segmentation is therefore a problem that still remains unsolved.

Berkeley's research of 2004, along with many other past projects and documents related to image segmentation constituted the basis of this project. From that point on, project extended existing research to blurred images, focusing on how the blurriness of an image affects both machine and human image segmentation.

The area of segmentation in image processing is extremely interesting, not only for its applications, but also for all the related methods and algorithms that have been developed and are being used. Moreover, it is a very important aspect for many sciences, specialised in each one. Depending on the use of the results of image segmentation, image analysis is specified around different baselines. Every case is different, a fact which extends the subject even more.

Segmentation is used to modify the representation of an image with the intention that its appearance becomes more significant. In the process, an image is analysed into meaningful pieces, always according to the purpose of processing the image. As said, this process is used in many sections of science and so it is of vital importance. Some of these sections include medical imaging, satellite images analysis, face recognition, fingerprint recognition, traffic control systems, brake light detection and machine vision [1].

As a vital section of image processing, image segmentation constitutes an extensive area of research in computer vision. There are many approaches to grey scale image segmentation, but nowadays most of the digital images are colour images [2].

Image segmentation has been applied on a variety of types of images. As it can be assumed by the applications of image segmentation previously referred to, some kinds of images where segmentation is applied are: medical images, portrait pictures, items pictures, sceneries, roads and highways and satellite images.

Machine image segmentation algorithms and methods have been developed in order to support human capability of analysing a picture and produce useful information. For example, a picture could be segmented in order to classify its objects or to categorise it as a whole. Image segmentation could also be applied by using different criteria, in order to supply scientists with a variety of information about the image. For instance, in segmenting medical images, machine image segmentation can use a variety of criteria to identify different characteristics and therefore provide broader knowledge. Such an example is shown in Figure 1.

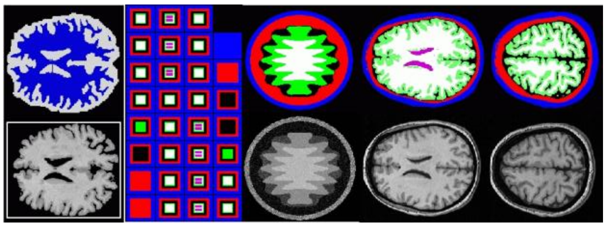


Figure 1. Brain Segmentation Medical Image [3]

All of the above applications and uses describe how important any improvement on machine image segmentation is. The smallest change in machine segmentors could cause a chain of changes in problem approach. A reason for the difference in performance between machine and human segmentation is the fact that machine segmentors enter a greater degree of detail, identifying as segments objects that humans would disregard in an image. This belief is what brought up the approach through blurred images.

1.1 Aim

The aim of this project was to evaluate the effects of blurred images on machine and human segmentation performance and to observe human cognition on these images with respect to image contents. In the same context, project aimed to provide ground truth from the dataset gathered from human segmentations. Based on approaches described in previous research, human segmentations were gathered along with machine segmentations and then comparison of the two was performed. Evaluation was executed using Precision Recall graphs framework, a technique successfully used in previous researches. Moreover, human segmentations gathered from experiments were observed in order to identify relations of image features in human cognition of images. Related past work and already implemented software were discussed in Literature Survey to provide a better understanding of the problem and identify all available resources. Design of experiments, implementation and evaluation procedures was the next part of the project, so as to decide on the most appropriate technique for every aspect of the project. Upon completion of design, implementation of any required written software was completed and then experiments took place. Results acquired from experiments were afterwards analysed, successfully leading to conclusions.

1.2 Objectives

Project's main objectives can be described by the two main hypotheses that constituted it:

Hypothesis I

"If blurred images are used in both machine and human segmentation, difference in their performance will be smaller than in the case of using high resolution images."

The approach to this hypothesis followed structure of [6], analysing data in the same way and resulting in conclusions again in the same manner. For the purposes of this hypothesis, human and machine segmentations were collected for a group of ten images, blurred at four levels, both in colour and greyscale form. When data were obtained, evaluation of results using the framework of Precision Recall graphs followed, again as done in [6]. Upon displaying results of segmentation both for human and machine segmentation, differences in the cases of original images and the blurriness levels became obvious. Based on the resulting PR graphs, difference in performance of machine and human segmentation was observed allowing for comparison between the four different blurriness levels.

Hypothesis 2:

"If people are asked to segment a blurred image where detail is less, they will vary the quality of their segmentation from a high resolution image to the blurred one in relation to the contents of the image."

In order to be able to test this hypothesis, the group of all human segmentations was saved for every image, both in colour and greyscale. Afterwards, comparisons were carried out, leading to observations of the differences between complex and simple images' segmentation. Testing of validity for this hypothesis was more of a visual procedure, rather than statistical. Segmentations of images were examined and related to each image's characteristics. Attention was given to the simplicity of the image regarding its colours' contrast, amount of different shapes and textures and the general complexity in the image.

Moreover, questionnaires were distributed to the participants. Purpose of these was to identify what made segmentation difficult for the participants and how it made them feel when they had to segment a complex picture in relation to the same process but on a simple picture.

1.3 Research

As described, this project extends on an existing research. Blurred images are introduced to image segmentation, and the difference in performance of humans and machines is evaluated. This notion extends the main topic of [6], using the Berkeley segmentor for this project's purposes. Performance of machines' and humans' segmentation on this type of images is unknown. While in [6] machine's performance is lower than that of humans, use of blurred images was expected to affect both machines and humans performance.

So as to be able to understand how the already implemented and available code works, how to use it and how to develop the project, extensive study on various areas of computer vision was required. Image segmentation and blurring constituted the main research, since these were the essential sections of the project. It was considered critical to study and understand already implemented and used segmentation algorithms, as well as blurring methods.

The second more important feature of this project was related to human cognition of images. There is no significant previous research on this subject, therefore this research objective stands alone, focusing on a more psychological rather than technical notion.

Both of the hypotheses tested in this project aimed to form a basis for further research.

Chapter 2

Literature Survey

Research in image segmentation goes back in time and includes a tremendous amount of papers and projects analysing its aspects. Its applications aim in acquiring essential information, therefore any improvement on the current knowledge would be appreciated.

As it is so important in the sector of science, image segmentation has been developed through the use of various algorithms, their performance comparison, their improvement and extension to include more image features. Extensive research also included comparison on how well machine image segmentation algorithms compare to human image segmentation. Despite the continuous progress, there still exists significant difference between the machine and human image segmentation. Most of human segmentation results agree with each other, and almost all of machine algorithms seem to perform as well as each other too [6].

Each of the various algorithms differs regarding its focus on the image features that define its analysis characteristics. Again, this depends on the purpose of the algorithm (e.g. applied on grey scale or colour images). A technique of assessing how well an algorithm performs is by producing a “Precision-Recall” graph.

All of the above relate to the current project as the main objective of this project was to analyse the effects of using blurred images in both machine and human segmentation. Basically, this project can be considered as an extension to Malik et al. research for University of Berkeley [6]. Implementation and research of the current project used the same resources to the above resource with the intention of extending its notions to blurred images.

This section of the document explores all main topics related to the project in detail and discusses all of the technical aspects of the system such as the implementation of image blurring and how a Precision-Recall graph can be produced and interpreted.

2.1 Image Segmentation

2.1.1 Definition

"Segmentation is the process of finding a connected region within the image with a specific property such as colour or intensity, or a relationship between pixels, that is, a pattern." [8]

Based on the purpose of processing an image and segmenting it, characteristics of segmentation procedure can vary. These differences may include the variety of features of an image being studied, e.g. identified a special need on texture. In addition, the required level of detail in segmenting an image can potentially appear to be different between segmentation processes, e.g. medical image analysis in comparison to traffic lights mechanism.

2.1.2 Human – Machine Segmentation Comparison

The idea behind this comparison is to test whether and how efficiently machine algorithms segment an image into meaningful pieces the same way humans do. Over the years, the number of detailed image characteristics being used in machine segmentation has increased. More features of an image are nowadays used to describe it, making machine segmentation more efficient.

As components contributing to the performance of a machine segmentation algorithm that are taken into consideration are now more than they used to be, algorithms are naturally more efficient. They appear to segment images in a more detailed manner and in a more efficient way compared to humans' segmentation than they used to.

The results produced by the two types of segmentation have a significant difference, but this distance is gradually getting smaller as better segmentation algorithms are being developed. What this project observed is the comparison of humans-machine segmentation on high resolution images against segmentation on blurred images.

The main point is that blurred images do not reveal so much detail in an image, as they hide the real edges of an object.

2.2 Image Features of Boundary Detection

As previously said, segmentation can be based on a variety of image characteristics. This section shall examine and discuss some of these image features. Information provided by these features concerning the image shall be presented, as well as the benefits of their use.

As a reminder, the project was based on work of Malik et. al. [6]. The image features used are therefore the same as those used in that specific research project. In that particular project, the research team implemented a new improved segmentor, now called Berkeley's Segmentation Engine (BSE), which appears to be the best performing segmentor until now. You might recall that BSE is also the segmentor used in the current project.

The next few sections shall explain the image characteristics BSE uses. These include four gradient-based features, more specifically two brightness features, the oriented energy and the brightness gradient, one colour feature, the colour gradient and one texture feature, the texture gradient. However, it is considered essential to explain what gradient-based features are before moving to the explanation of the boundary detection image features.

2.2.1 Gradient-Based Features

A gradient-based paradigm is introduced for more complex features. The procedure of its analysis is the following.

At a point (x, y) in image, a circle is drawn with radius r and divided along the diameter at orientationθ. If the pixels lie in different segments, then it is expected to find discontinuity somewhere along the line [9]. If no such discontinuity is encountered, then the similarity between the pixels should be large.

Gradient function $G(x, y, \theta, r)$ compares the contents of the two resulting disc halves. If a large difference between the disc halves is identified, a discontinuity in the image is present along the disc's diameter. [6]

2.2.2 Oriented Energy

Oriented energy (OE) is a standard mean of detecting brightness edges in images. Based on the difference of the amount of brightness at particular points in an image, we can identify boundaries based on brightness. In other words, points where brightness appears to differ from the rest of the image either by being less or more, can be declared as belonging to a brightness boundary, as it clearly separates levels of brightness in the image.

Based on Morrone and Burr [10] approach for detecting edges and lines in an image using their local Fourier representation, the detection and localization of these composite edges can be calculated using OE. OE is defined as:

$$OE_{\theta,\sigma} = (I * f^{e}_{\theta,\sigma})^2 + (I * f^{0}_{\theta,\sigma})^2 \, . \tag{2.1}$$

Where $f^{e}_{\theta,\sigma}$ and $f^{0}_{\theta,\sigma}$ are a quadrature pair of even and odd-symmetric filters at orientation θ and scale σ.

2.2.3 Colour Gradient

Analysis of difference between colour distributions on a set of pixels can be done using two different techniques.

The first technique is based on density estimation using histograms. QBIC and Blobworld image retrieval systems use three-dimensional colour histograms as region features. Blobworld smoothes histograms to prevent aliasing of similar colours. QBIC on the other hand models the perceptual distance between bins explicitly.

The second approach to colour distributions avoids quantization artefacts by using Mallows [11] or Earth Mover's distance (EMD) [12] to compare colour distributions.

However, in BSE implementation, a technique of modelling the colour distribution with accuracy according to human perception was desired, without that affecting the computational feasibility. Therefore, BSE is "based on binning kernel density estimates of the colour distribution in CIELAB using a Gaussian kernel, and comparing histograms with the χ^2 difference." [6]

The χ^2 histogram difference does not make use of the perceptual distance between bin centres. Without smoothing, perceptual similar colours can therefore produce disproportionately large χ^2 differences.

Distance between points in CIELAB space is perceptually meaningful in a local neighbourhood. As a result, binning a kernel density estimate whose kernel bandwidth σ matches the scale of this neighbourhood means that perceptually similar colours will have similar histogram contributions. Beyond this scale, where colour differences are perceptually incommensurate, χ^2 will regard them as equally different. In [6] Malik et al. supported that this combination of a kernel density estimate in CIELAB with the χ^2 histogram difference is a good match to the structure of human colour perception.

2.2.4 Brightness Gradient

In the analysis of the brightness gradient, histograms of L* values are computed. Additional challenges for density estimation are presented through the colour gradient. This is because the pixel values are in the 2D space. When using 2D kernels and 2D histograms one typically reduces both the number of kernel samples and the number of bins in order to keep the computational cost reasonable. However, this compromises the quality of the density estimate.

Rather than compute the joint gradientCG^{ab}, marginal colour gradients for a* and b* are computed and the full colour gradient is taken to be the sum of the corresponding marginal gradients:

$$CG^{a+b} = CG^{a} + CG^{b} \; . \tag{2.2}$$

This is motivated by the fact that a* and b* channels correspond to the perceptually orthogonal red-green and yellow-blue colour opponents found in the human visual system.[13]

2.2.5 Texture Gradient

The last feature necessary for edge detection process is texture gradient. This can be calculated using a formulated directional operator, which measures the degree to which texture of scale r varies at an image location (x, y) in direction θ. Similarly to brightness and colour gradients, attention is given to potential discontinuity in the two halves of a disc centred on a point and divided in two along a diameter. Such discontinuity indicates texture dissimilarity.

In the process of texture processing, a filter bank is used as shown in Figure 2(a) below.

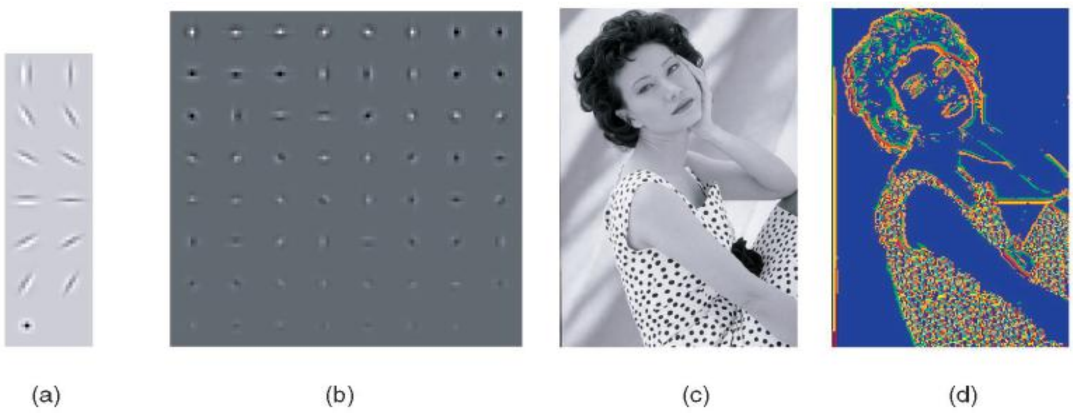


Figure 2. Computing Textons. (a) Filter Bank: The 13-element filter bank used for computing textons. (b) Universal Textons (c) Image and (d) Texton Map [6]

Six pairs of lengthened, oriented filters are used, as well as a centre-surround filter. Oriented filters are in even/odd quadrature pairs, and are the same filters used to compute oriented energy. [10]

However, unlike oriented energy, even and odd filter responses are not united. For each pixel, a vector of thirteen filter responses is associated to it. Filter responses are not contrast-normalised, unlike [14], as this type of normalisation does not improve performance, since it amplifies noise more than signal. Each disc half contains a set of filter response vectors which can be visualised as a cloud of points in a feature space with dimensionality equal to the number of filters.

There are several possible approaches for the measurement of filter response vectors. One way is the evaluation of a wide range of texture descriptors in this framework and examine several of their characteristics e.g. how should the distributions of filter responses be compared [15]. However, the chosen approach used in BSE is different. It follows the approach described in [14], which introduced the use of textons. The notion of textons shall now be explained.

Textons Definition

Research with texture pairs having identical second-order statistics has revealed that the pre-attentive texture discrimination system cannot globally process third- and higher-order statistics, and that discrimination is the result of a few local conspicuous features, called textons.[16]

Outputs of linear filters responses are used in the texture gradient context and considered as points in a high dimensional space. Vector quantization is performed in this high dimensional space to find prototypes. From there, these prototypes are used as textons.

When textons have been acquired, a universal texton vocabulary can be constructed by processing a large number of natural images. Alternatively, it is possible to find textons adaptively in windows of images. The optimal number of textons k depends on the choice between universal and image-specific as well as the scale r of the texture gradient operator and the size of the image.

K-means are used for filter response vectors clustering. The process of texton mapping follows, in which each pixel is mapped to the texton nearest to its vector of filter responses as shown in Figure 2 (c) and (d) above. Based on this process, the image can be analysed into texton channels, each of which is a point set.

Texton Histograms Computation

When the texton map is completed, texton histograms can be computed. For producing them, hard binning without smoothing is used. Soft binning could also be used, but is computationally expensive. Since adjacent pixels have correlated filter responses due to the spatial extent of the filters, hard binning is not a problem. Consequently, data is already somewhat smoothed, and pixels in a disc are likely to cover fewer bins ensuring more samples per bin.

2.2.6 Localisation

Since each pixel is classified independently, spatially extended features are problematic for a classifier, as both on-boundary pixels and nearby off-boundary pixels will have large gradient values. The texture gradient (TG) is particularly prone to this effect due to its large support. The bands of textons present along such edges often produce a larger TG response on each side of the edge than directly on the edge.

Non maxima suppression is typically used to narrow extended responses, but multiple detections require a more general solution. The symmetric nature of the texture gradient response is exploited to localise the edge accurately and eliminate the double detections.

Process of making the spatial structure of boundaries available to the classifier will now be described. Firstly, raw feature signals are transformed in order to emphasize the local maxima in a manner that simultaneously smoothes out multiple detections. Given a feature f(x) defined over spatial coordinate x orthogonal to the edge orientation, consider the derived feature $\hat{f}(x) = f(x)/d(x)$, where $d(x) = -|f'(x)|/f''(x)$ is the first order approximation of the distance to the nearest maximum of $f(x)$. The smoothed and stabilised version is

$$\hat{f}(x) = \tilde{f}(x) \cdot \left(\frac{-f''(x)}{|f'(x)| + \epsilon}\right). \tag{2.3}$$

with ϵ chosen to optimise the performance of the feature. Factor $\tilde{f}(x)$ is a smoothed estimate of the underlying gradient signal that eliminates the double peaks.

A cylindrical parabola over a 2D circular window of radius r centred at each pixel is fitted to robustly estimate the directional derivatives and the smoothed signal. The axis of the parabolic cylinder is constrained to lay parallel to the image plane and encodes the edge location and orientation. The height encodes the edge intensity; and the curvature of the parabola encodes localisation uncertainty. Data points are projected inside the circular fit window onto the plane orthogonal to both the image plane and the edge orientation, so that the fit is performed on a 1D function.

The least squares parabolic fit ax2 + bx + x provides directly the signal derivatives as f''(x)=2a and f'(x) = b, as well as $\hat{f}(x) = c$. Thus, the localisation function becomes $\hat{f} = -2c^{+}a^{+})/(|c| + \epsilon)$, where c and a are half wave rectified. This rectification is required to avoid nonsensical sign changes in the signal when c and a are multiplied together.

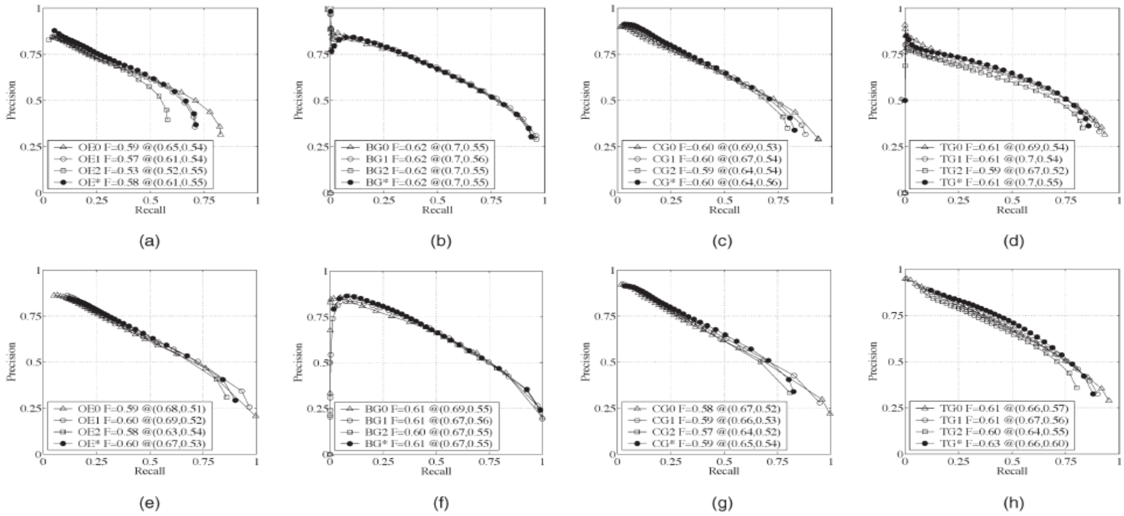


Figure 3. (a) Raw OE, (b) Raw BG, (c) Raw CG, (d) Raw TG, (e) Localized OE, (f) Localized BG, (g) Localized CG, (h) Localized TG

This completes the explanation of the whole of image features used in BSE along with their importance. BSE combines all of the cues into a single function $P_b(x, y, \theta)$ which gives the posterior probability of a boundary appearing at each pixel (x, y) and orientation θ.

When the boundaries are identified, the next step in a segmentor is judging the quality of a boundary detector. This is essential in order to optimise the parameters of the segmentor and compare it to those of others. A methodology for doing so is covered by the precision-recall framework. This shall now be explained in the next section of this document.

2.3 Evaluation Methodology

Boundary detection is formulated as a classification problem of discriminating non-boundary from boundary pixels. After that, Precision-Recall framework is applied using human marked boundaries as ground truth.

2.3.1 Precision-Recall Curve

"A parametric curve that captures the trade off between the accuracy and noise as the detector threshold varies, a rich descriptor of performance [6]. It is a standard evaluation technique in the information retrieval community [17]."

A decision made by a classifier can be represented in a structure known as a confusion matrix or contingency table. This confusion matrix is separated in four categories: True Positives, False Positives, True Negatives and False Negatives. In precision recall (PR) space, Recall is plotted on the x-axis and Precision on the y-axis. Recall can be considered as the True Positives Rate, i.e. the fraction of true positives that are detected rather than missed. Precision on the other hand measures that fraction of examples classified as positive that are indeed positive. [18]

PR measures are particularly meaningful in the context of boundary detection when considering applications that make use of boundary maps. It is sensible to characterise higher level processing in terms of how much true signal is required to succeed Recall, and how much noise can be tolerated in Precision.

A particular application can define a relative cost α between these quantities, which focuses attention at a specific point on the precision-recall curve. The F-measure [17] is defined as:

$$\mathrm{F} = \mathrm{PR}\,/(\mathrm{aR} + (1 - \mathrm{a})\mathrm{P}) \quad . \tag{2.4}$$

It captures the trade off as weighted harmonic mean of P and R. The location of the maximum F-measure along the curve provides the optimal detector threshold for the application given α.

Precision-Recall Measures

In order to compute precision and recall, it must be determined which true positives are correctly detected, and which detections are false. Each point in the precision-recall curve is computed from the detector's output at a particular threshold. Additionally, binary boundary maps are available as ground truth from the human subjects.

At this point, the matter of computation of precision and recall of a single thresholded machine boundary map should be done given a single human boundary map. One approach of doing so is to correspond coincident boundary pixels and declare all unmatched pixels either false positives or misses. However, this approach does not tolerate any localisation error and would therefore over-penalise algorithms that generate usable, however slightly mislocalised boundaries.

A second approach proposed by D. Martin, C. Fowlkes, and J. Malik in [19], describes the addition of a modicum of slop to the rigid correspondence procedure described above in order to permit small localisation errors at the cost of permitting multiple detections. However, an explicit correspondence of machine and human boundary pixels is the only way to robustly count the hits, misses, and false positives that we need to compute precision and recall. It is important to compute the correspondence explicitly in order to penalise multiple detections.

Nevertheless, in BSE implementation, computing the precision and recall for a single human segmentation while permitting a controlled amount of localisation error was provided by the correspondence computation [6]. Joining humans' boundary maps is not effective because of the localisation errors present in the data set itself. A proper approach to combining human boundary maps would require additional correspondences or even estimating models of the humans' detection and localisation error processes along with the hidden true signal.

In order to resolve these issues, machine boundary map is corresponded separately with each human map in turn. Only those machine boundary pixels that match no human boundary are counted as false positives. The hit rate is simply averaged over the different humans, so that to achieve perfect recall the machine boundary map must explain all of the human data. The intention of this approach was to estimate precision and recall to match as closely as possible the intuitions one would have if scoring the outputs visually.

Figure 4 presents the performance of Berkeley's boundary detector compared to classical boundary detection methods and to the human subjects' performance. A precision-recall curve is shown for each boundary detector: 1) Gaussian derivative (GD), 2) Gaussian derivative with hysteresis thresholding (GD+H) 3) A detector based on the second moment matrix (2MM), 4) BSE gray-scale detector that combines brightness and texture (BG+TG), and 5) BSE detector that combines brightness, colour, and texture (BG+CG+TG).

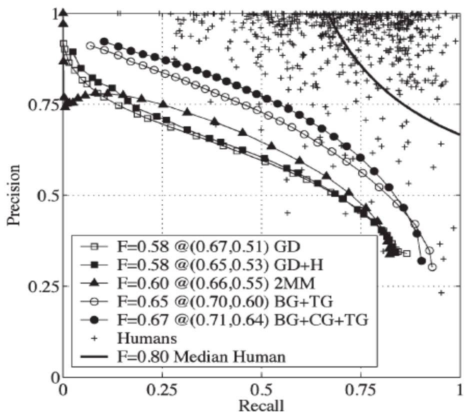


Figure 4. Two Decades of Boundary Detection

2.4 Cue Optimisation

In order to accomplish the best results, it is best to optimise each cue individually before combining all of them into a single detector. This was achieved in BSE by applying coordinate ascent on each cue's parameters with the intention of reaching high precision and recall. The result provides optimised cues with respect to the ground truth data set so that no change in any single parameter improves performance.

The four image features used (oriented energy OE, brightness gradient BG, colour gradient CG and texture gradient TG), all have a scale parameter. For OE the scale σ is the bandwidth of the quadrature filter pair, while for the rest cues scale r is the radius of the disc. For each cue, the optimal one octave range for each cue is determined: in units of the image diagonal, the ranges are 1.4 to 2.8 percent for OE, CG ant TG, and 0.75 to 1.5 percent for BG. Middle scale always performs best, except in the case of raw OE where the largest scale is superior.

Based on the PR graphs for each cue at optimal scales both displaying with and without applied localisation presented in [3], it can be assumed that localisation is not required for BG and CG, but helpful for both OE and TG. Graphs can be viewed in Figure 3 above.

Localisation may lead to two potential benefits, narrowing peaks in the signal and merging multiple detections. Scale of OE is rather large so localisation is effective at narrowing the wide response. TG suffers from both multiple detections and a wide response, both of which are ameliorated by the localisation procedure.

Optimisation of the kernel size used in the density estimation computations for BG and CG can be viewed in Figure 5 below. For these cues, distributions of pixel values in two half discs are compared to see if they are brightness (L*) or colour (a*b*).

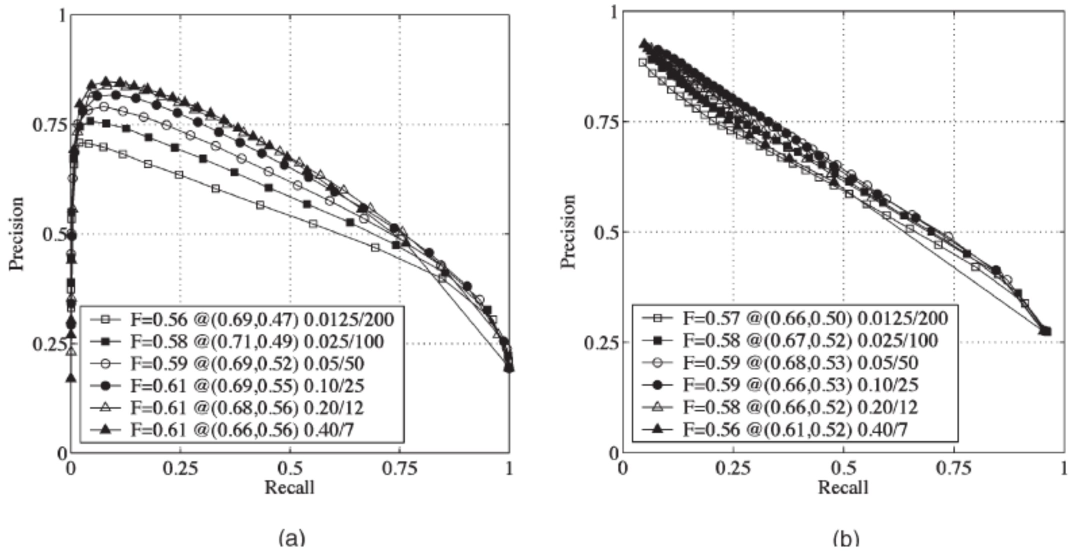


Figure 5. Brightness gradient, (b) Colour gradient, Kernel bandwidth for BG and CG kernel density estimates

As described in [6], the difference between separated version of colour gradient CG^{a+b} and joint version CG^{ab} is minimal. The joint technique though is more computationally expensive due to the additional dimension in the kernels and histograms. If computational expense is kept constant, then the marginal method is superior because of the higher resolution afforded in the density estimate. In every case, the marginal approach to computing the colour gradient is preferable.

Texture gradient on the other hand has some additional parameters beyond r and $\in$ to tune, related to the texture representation and comparison. As previously analysed, texture gradient TG is used to quantify the difference in the distribution of filter responses in the two disc halves. The same even and odd symmetric filters that are used for oriented energy are also used for these filters with six orientations along with a centre-surround DOG [15]. Based on what Levina [20] supported, a single-scale filter bank at the smallest scale was preferred for the BSE implementation.

Regarding distribution estimation issues, the texton approach of Malik et al. [14] is used in BSE, an approach which estimates the joint distribution with adaptive bins by clustering the filter responses with k-means, and compares histograms using the χ^2 measure. As for the type of the textons used, computational cost of either using image-specific textons or universal textons is approximately equal. However, image-specific textons were chosen over universal textons due to their convenience, even though universal textons are more appealing as they can be used to characterise textures in an image-independent manner.

Results from both image-specific and universal textons are comparable, as the image-specific textons vary between training and testing images. Since the texture gradient is only dependent on having good estimates of the distribution in each half-disc, the identity of individual textons is unimportant.

2.5 Cue Combination

Completing each cue’s optimisation, the process of combining all cues into a single detector can then follow. This process is handled as a supervised learning problem, where the combination rules are learned from the ground truth data.

Logistic regression is used for cue combination. Firstly, whether any of the cues is redundant given the others is determined. In fact, both OE and BG detect discontinuities in brightness. It appears that BG is superior to OE, either used in conjunction with the TG alone, or with TG and CG together. Therefore, OE can be dropped from the essential cues.

Moreover, each cue can be calculated at multiple scales, however, only TG contains significant independent information at the different scales. Even so, multiple TG scales are not beneficial in the case of TG being combined with other cues. In some cases, e.g. using TG in combination with BG and CG, multiple scales do not improve overall performance. In some cases performance can even degrade as additional scales may introduce more noise than signal. Therefore, to serve simplicity as well as performance, only the middle scale of each feature shall be retained.

2.5.1 Classifiers

Another characteristic of BSE that was tested was whether selecting the logistic model to generate results was a good choice. In order to perform the test, a number of classifiers was compared, each of which was trained on the human segmentation data set. All five classifiers used for the test are listed below:

- **Density Estimation**

Density estimation is done with adaptive bins provided by vector quantisation using k-means. Each k-means centroid provides the density estimate of its Voronoi cell as the fraction of on-boundary samples in the cell.

- **Classification Trees**

Domain is partitioned hierarchically with top-down axis-parallel splits. When a cell is split, it is split in half along a single dimension. Cells are split greedily so as to maximise the information gained at each step. The effect of this heuristic is to split nodes so that the two classes become separated as much as possible.

- **Logistic Regression**

In this technique, initialisation is random, and convergence is fast and reliable by maximising the likelihood with about five Newton-Raphson iterations. Two variants are also considered: quadratic combinations of features and boosting using the confidence-rated generalisation of AdaBoost by Schapire and Singer. [21]

- **Hierarchical Mixtures of Experts (HME)**

This model was described by Jordan and Jacobs [22] and is a mixture model where both the experts at the leaves and the internal nodes that compose the gating network are logistic functions. Model is initialised in a greedy, top-down manner and fit with EM.

- **Support Vector Machines**

SVM package libsvm [23] is used to perform a soft-margin classification using Gaussian kernels.

From the results of Malik et al. in [4], it is shown that segmentations of a single image by the different subjects are highly consistent, therefore all human-marked boundaries are considered valid. For training, an image location (x, y, θ) is declared to be on-boundary if it is within $\Delta x \leq \sqrt{8}$ pixels and $\Delta\theta = 30°$ degrees of any human-marked boundary. The remainder is labelled off-boundary.

This classification task is characterised by relatively low dimension, a large amount of data, poor class separability, and a 10:1 class ratio. The maximum feasible amount of data, uniformly sampled, is given to each classifier. A high degree of class overlap in any low level feature space is inevitable because the human subjects make use of both global constraints and high-level information to resolve locally ambiguous boundaries.

Based on the above, conclusions regarding the efficiency of use of logistic model were drawn for BSE. Overall, the performance of all classifiers is approximately equal, and logistic model is preferred due to performance, model complexity and computational cost.

2.6 Programming Background

As the mathematical analysis and explanation is now complete, we can move on to explanation of the actual implementation of the project. The programming aspects of the project can be broken into the following components: Blurring Images Mechanism, Machine Segmentor, Human Segmentation Support Interface, and Results Presentation Interface.

2.6.1 Blurring Images Mechanism

As this project is mainly based on blurred images, the ability of blurring a picture was essential. This was performed for every image in the test data set for four levels of blurriness.

Image blurring process is actually a procedure that suppresses noise and small fluctuations [26]. In the frequency domain, this process refers to the suppression of high frequencies. A smoothing filter was built in Matlab by using function fspecial (special filters). A Gaussian function was used to calculate the mask. The standard deviation (sigma) and the size of the masks could be defined by the user, a useful tool for changing the blurriness level. If sigma and mask size are defined small, then the filter shall only blur the image slightly. In order to blur an image heavily, larger sigma and mask size shall be applied.

The procedure of blurring was performed as follows. Every image being tested was blurred on various levels. Therefore, the amount of blurriness was incremented gradually and it the resulting image was stored.

2.6.2 Machine Segmentor

This project does not include the original implementation of segmentors or classifiers. Already implemented and tested mechanisms will be used. In particular, the project undertakes the work of Malik et al. in [5] [7], using the open source Matlab code produced for BSE. In [6] four boundary detectors were tested against the one being implemented, verifying that the resulting segmentor (BSE) had the best performance. However, this project does not focus on how well each algorithm performs against the rest of them. The main consideration is the comparison of machine segmentation to actual human segmentation overall. Intention of the project is to study how blurred images affect the performance of both human and machine segmentation. Therefore, BSE boundary detector, the best one developed, namely BG-CG-TG from the gradients it uses, was the boundary detector used.

2.6.3 Human Segmentation Support Interface

About performing the human segmentation experiments, an interface for handling this was required. At first, implementing such an interface in Matlab was intended, but it was proved that a complete Java application was already available from [4] [7]. However, a graphical user interface was planned to use Matlab's GUI support to connect machine segmentation with human segmentation. This interface was intended to be used in the experimental phase of the project, aiming at both experienced and non-experienced users. This implied a necessity for simplicity and ease of use in the graphical user interface. User interface design would follow the basic rules regarding both design and usability principles. [24] [25]

The mechanism supporting human segmentation was retrieved from the already implemented system from Malik et al. in [6]. It would be encoded into a carefully designed interface, which would be assessed and reformed to meet project’s needs regarding blurred images.

2.6.4 Results Presentation Interface

When experimentation phase is complete and all the resulting data have been gathered, next step shall be observation and conclusions. This aspect of the system shall be responsible for presenting the resulting Precision-Recall graphs and the statistical analysis of the results.

Again, the principles of User Interface design and usability will guide the interface implementation. As this interface will be responsible for presenting both tables and graphs, attention shall be given to provide the user with options on how to view the results. Further analysis regarding this area of the project is included below, in section 2.6.6, where approach to research users interfaces characteristics is studied.

2.6.5 Graphical User Interfaces using Matlab

As previously described, main implementation of the project shall be done using Matlab. Since Matlab is mostly a mathematical rather than user supportive environment, research on potential tools regarding graphical user interfaces provided by Matlab was considered essential. Research showed that Matlab actually supports graphical user interfaces in a very good manner, without of course providing detailed handling of interfaces design.

However, properties available are enough, as the project does not aim to a special audience, e.g. children. From that point, accessibility was certainly considered, but a more detailed design, e.g. colour handling did not seem essential. Therefore, Matlab graphical user interface design tools appeared to be more than enough.

In this section of the project, guidelines towards interface design were [27] and [28].

2.6.6 Graphical User Interfaces for Research

This project intended to be used for research purposes to extend the current knowledge. This means that researchers belong to the group of primary users of the system implemented. Therefore, their needs are essential for the project and needed to be taken into consideration.

Apparently, this notion is new and has no background research. No available knowledge regarding particular characteristics of graphical user interfaces that aim to be used into research is available.

In order to gather information concerning this subject, interviews and questionnaires shall be used in the requirements gathering. Current researchers both experienced and non-experienced computer users shall be included. In the requirements gathering process, project aimed to contact senior experienced researchers regarding this same subject in order to gather redesign suggestions and ideas.

2.7 Expectations

At the initial states of the project, still in the process of research, expectations of the project were also considered. The first prospects regarding the performance of machine algorithms and humans in segmentation were that the difference between them would decrease when dealing with blurred images. The median of human segmentations was expected to appear closer to the machine segmentors, as segmentation on a less detailed image seems to be more difficult for humans than that of a natural image.

2.8 Summary

In this section, the background knowledge necessary for the project, as well as resources and mechanisms were analysed, discussed and related to the project's aims.

In the first section segmentation was defined and its appeal to the project was analysed. We then moved on to discussion of image features processed and used in the segmentor used, commenting on their use and various approaches of handling them.

When this was done, discussion for evaluation methodology and results collecting took place. In this sector importance of precision-recall graphs was analysed, along with their interpretation.

Continuing the analysis, cues' optimisation was described, compared their value to the project and explained their appropriate parameters and level of detail. When optimisation discussion was complete, we moved on to cue combination to explain how all the import features from images could be used to create one single detector. Discussion on classifiers used in [6] followed next, including commenting on the techniques.

The most crucial section of the project followed, unfolding the aspects of programming of the project. Underlined was the characteristic of the use of Matlab for every feature of the system. Moreover, it was discussed that the segmentation mechanism to be used shall partly consist of the system implemented for BSE in [5]. From that the current system shall extend to cover blurring mechanism and interfaces for user support.

Analysing the graphical user interfaces in the programming sector of this document, attention was given to general design rules and usability principles. At this state, it was identified that Matlab provides all the essential features for implementing a user interface.

Another feature noticed at the point of user interface design needs was the need for further study on designing interfaces especially for research purposes, particularly because there is no previous research regarding the subject.

Summarising, this project aims to provide knowledge regarding the effects of using blurred images in the variation of behaviour between human and machine segmentation. Using Malik et al. research for University of Berkeley as guideline, project intends to extend the current level of knowledge and provide new information regarding image segmentation from both machines and humans.

Chapter 3

Design of Experiments

Before beginning the detailed analysis of how each of the experiments was designed to be executed, an outline of the order of each experiment and what each one contained will be presented.

As described in the introductory chapter, two hypotheses determined the project. Both of these hypotheses required human segmentation results to proceed into validity check. Therefore, the first experiment that was carried out was that of human segmentation by all the participants of the project. In order however to perform this experiment, blurred test images had to be created and provided to the participants.

This is why the initial phase of experiments included the blurring of a collection of ten representative 481x321 RGB images from the Corel image dataset. These images were selected from a sample group of the Corel image dataset, the same one that was used in [6] as well. When selected, the collection of the images was blurred and saved. Images were blurred at four different levels, one level of zero blurriness (original), one of maximum blurriness and two medium levels, before being stored. Procedure of blurring was executed both on colour and greyscale images, providing the participants with a total of eighty images.

When the collection of blurred images was completed, human segmentation experiments could be performed. This was a crucial part of the project, as participants would provide all the necessary ground truth of human segmentations that would be used in hypotheses validity testing. Using already implemented code developed for the purposes of [4], a java application was used by every participant in order to complete the human segmentations experiment.

While participants were completing the human segmentations, machine segmentation could be performed. This was developed on its own at first, but was then integrated with the final stage of machine and human segmentation evaluation. An open source software package provided by University of Berkeley, called "Berkeley Segmentation Dataset and Benchmark" (BSDS), implemented in Matlab and available from the Internet [7], provided all the necessary functionalities to perform both the machine segmentation and the evaluation of performance using PR graphs framework. Machine segmentations were also stored once they were gained. The same happened to resulting PR graphs from the evaluation experiment.

For testing the second hypothesis of this project, collection of human segmentations was sufficient. Hypothesis was tested for validity by viewing the human segmentation on top of the image to which they corresponded. This functionality was already available from the software package used to perform the human segmentation experiments; therefore resulting in analysis for the second hypothesis was easier than the procedure necessary for doing so for the first one. Remarks collected from the above procedure and participants' answers in the questionnaires that they were distributed, formed enough information for testing the second hypothesis validity.

Having all necessary data gathered, all experiments could be efficiently completed and provide results that would later lead to conclusions. This completes the outline of the structure of experiments, starting from producing the required data and finishing off with analysis. The next sections of this chapter shall now describe in detail each of the components of experiments' development.

3.1. Blurring Mechanism

In order to perform the blurring procedure on eighty images in total (ten images in four different levels of blurriness, both colour and greyscale), software for doing so was implemented in Matlab (see Appendix B). Images were read in code as matrices of type double and then blurred using a rotationally symmetric Gaussian low-pass filter.

At first the three levels of blurriness were produced by changing the value of standard deviation sigma, but this was proved to provide insufficient variance in the successive blurriness levels. An example of the results of this approach can be seen in Figure 6 below.

(a) (b) (c) (d)

Figure 6. (a) Original Image, (b) Blurred Image with σ=8, (c) Blurred Image with σ=16, (d) Blurred Image with σ=32

As one can observe from Figure 6, blurring the images with the previously described method does not provide significant difference between the resulting images. Therefore, a different approach was favoured over that.

Again, a symmetric Gaussian low-pass filter was used, only this time the three different levels of blurriness were produced by changing the number of rows and columns of the filter. The standard deviation was kept constant at a value of 8, while the size of the filter was increased by 5 for each increasing level of blurriness. Lower blurriness used a filter of square matrix of size 5, while the maximum one used a filter of square matrix of size 15. The resulting images are displayed in Figure 7.

(a) (b) (c) (d)

Figure 7. (a) Original Image, (b) Blurred image with matrix size 5, (c) Blurred image with matrix size 10, (d) Blurred image with matrix size 15

Clearly, the second approach to the blurring procedure results in more significant difference between the levels of blurriness. This is the reason this approach was finally selected for blurring the test images rather than the first approach described.

3.2. Human Segmentation Mechanism

As said before, this experiment was the most demanding of all, as well as the most important. It had the longest duration from all other experiments of the project, as it implicated both software use and interaction with users. In the process of setting up the java application and for general experiment guidelines on performing the human segmentation, research of [4] was used as the main resource.

Human segmentation mechanism is available online at [7], from where it was also downloaded for the purposes of this project. As stated in [4], the instructions given to the participants had to be intentionally vague in order to guide them to break up the image into a natural manner, while do not directing them on how to think about doing so. The instructions given to the participants were the following:

“Divide each image into pieces, where each piece represents a distinguished thing in the image. It is important that all of the pieces have approximately equal importance. The number of things in each image is up to you. Something between 2 and 20 should be reasonable for any of the images.”[4]

The document used [4] as guideline for this experiment was proved to be an excellent resource on how to perform the experiments. A note that should be made here is the fact that human segmentation software was run in mode hermit, i.e. it was run with access to images being in the local storage medium and not on an online server as the default version of the java application denotes.

This experiment was designed with attention to supervision and support to the participants. As this experiment was the one to provide the vital information of human segmentations, experiment was designed to include special care and support to each participant. Each participant shall be approached individually by the investigator, who would provide a ready setup environment for the participant to work in. All data would be stored on one machine. Continuous support and physical supervision would take place, with the investigator being fully concentrated on the participant, ready to answer potential questions, or note any comments and/or remarks of the participant.

3.3. Machine Segmentation Mechanism

For the second experiment, that of machine segmentation, already implemented Matlab software system BSDS available at [7] was used on Linux. Procedure of machine segmentation was implemented in integration to the performance evaluation procedure, as described previously. Here, it is worth reminding that the segmentor used in this experiment, as explained in Literature Survey, is the globally recognised and appreciated as the most efficient segmentor up to now Berkeley Segmentation machine (BSE). Almost every researcher uses that segmentor to perform image segmentations. It can therefore be said that the software used in the context of this experiment is trustworthy open source software, supported by the academic community of Computer Vision.

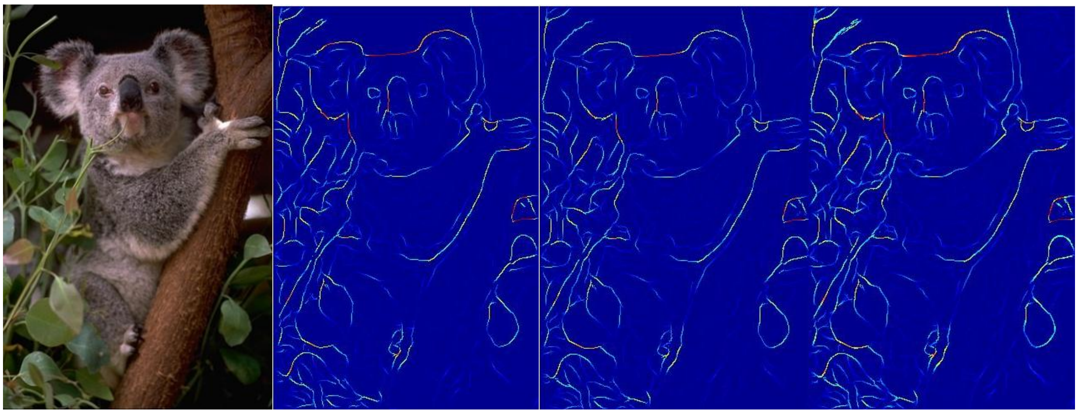
Original code acquired from [7] was used as provided. No changes were made to it. New software was implemented for making use of the functionalities available from the original software package. The developed code produced for performing this experiment by using ready functionalities has been included in Appendix B. Running the written code would request the image file the user desired to be machine segmented and then used in evaluation of performance. Once user had entered a valid filename for the image and the format of file desired (color/gray), machine segmentation was performed and presented to user along with the match between a random human segmentation and the produced machine segmentation. Finally, code computed and displayed the resulting PR graph for that particular image.

In order to gain efficient file access in the code, files were stored with numerical names for ease of access. Different folders were used for colour and greyscale images. Under the folder named after the number of participant to which the results belonged, segmentation file was named after its number (from 1 to 10), three intermediate zeros (000) and the number of blurriness in the image (0, 5, 10, 15). For example, a valid filename would be “800015”.

Software available from the BSDS, provided all the essential functionalities for the machine segmentation process, including reading image files, reading segmentation files from human segmentation experiment and transforming files from one type to another, for example from human segmentation boundaries to boundary probability maps. Additionally, BSDS supported a collection of boundary detectors, which could be used to perform machine segmentation. This collection of boundary detectors included Gaussian derivative, Gaussian derivative with hysteresis thresholding, a detector based on the second moment matrix, the Berkeley Segmentation Engine (BSE)'s greyscale detector that combines brightness and texture gradients (BGTG), and the BSE colour detector that combines colour and texture gradients (CGTG).

A combination of two detectors was used in code in order to get the best result machine segmentation. In particular, BSE's greyscale detector was firstly applied, combining brightness and texture gradients to identify the boundaries. After that, BSE's colour detector combining brightness and colour gradients was applied. By using these two detectors, the final result combined all brightness, texture and colour based detected boundaries, providing the best result available.

An example of machine segmentation can be viewed in Figure 8. This figure displays all the stages of machine segmentation, demonstrating each of the two detectors' individual results and then the final one (d).



(a) (b) (c) (d)

Figure 8. Machine Segmentation on an image of zero blurriness: (a) The original image, (b) the probability boundary map from BGTG detector, (c) the probability boundary map from CGTG detector and (d) the final result from machine segmentation, providing a BGCGTG result

3.4. Segmentation Evaluation Procedure

The last procedure of the experiments was the evaluation of both machine and human segmentations on the different levels of blurriness. As previously said, the Berkeley Segmentation Dataset and Benchmark includes a number of functionalities to perform machine segmentation using different algorithms and techniques to evaluate their performance in relation to human segmentation results. Therefore, the same original code acquired from BSDS that was used to perform the machine segmentation included functionalities for performance comparison. This is why the written software (appendix B), was structured in a way were machine segmentation would firstly be obtained and then it would automatically be fed into performance evaluation.

Framework of Precision Recall graphs was used for performing evaluation of performance. The written software was structured to perform evaluation of performance for both human and machine segmentations and then move on to the representation of results in the form of a final Precision Recall graph. This procedure was repeated four times for each of the test images, changing the level of blurriness for each time. The difference between human and machine performance was clearly visible in each case through the resulting PR graph. All resulting graphs were stored and then compared against each other, in order to observe variations of performance when changing the blurriness level.

Following exactly the same approach as Berkeley research, this experiment was guaranteed to succeed. Through understanding the system of BSDS and through the use of it, evaluation procedure was performed in the most sufficient and efficient way.

3.5. Research Ethics

In the duration of the project, a great amount of weight was given to Research Ethics [30] [31]. Standards of acceptable practice in the conduct of research that involves people were studied and analysed in order to perform experiments in the most appropriate manner. The ten standards of Nuremberg Code of 1947 [32] were studied to understand how to obtain a balance between expected benefits against the risks to those involved. Key principles of Human Research were also given attention, including minimizing risks to participants, researcher’s responsibility for correction of any negative effects, base participation on Informed Consent, use of consent forms, allowance of discontinuation at any time, provision of full debrief and encouragement of questions.

An important factor that should be referred to here is risk of performing the experiments implicated in the project. After discussing the form of experiments with a professional oculist, it was verified that experiments did not include any danger to participants. The only side effect that participants might have during the experiments was tiredness of their eyes. This was harmful, as it could not lead to a serious effect to the eyes, due to the short duration of the experiment.

This project was checked against the University of Bath Department of Computer Science policy for Research ethics [30]. A 13-point checklist was completed and can be seen in Appendix A. Additionally, all of the participants’ signed consent forms are attached at the end of the checklist.

Chapter 4

Experimental Results

In this chapter, we shall review the outcomes of the experiments described in the previous chapter. Results obtained shall be presented but not analysed in depth. Analysis shall take place in chapter 5, leading to conclusions.

4.1. Human Segmentations

The desired ground truth dataset regarding human marked segmentations was successfully acquired. Using the supporting functionalities of Java application deployed for this experiment, investigator was able to review the gathered data on top of the image to which they corresponded. Some of the results gathered in the human segmentation procedure can be viewed in Figure 9 below. Figure presents segmentations for the two different formats of images, both greyscale and colour, for all four levels of blurriness.

All human segmentations gathered from the experiments were successful. All participants provided a complete selection of segmentations, for all images, at all levels of blurriness, in both colour and greyscale. Acquiring eighty segmentation files from each participant, for a total of ten participants, provided a dataset of sufficient length to both be used in further research and for testing the validity of the hypotheses that describe the objectives of this project.

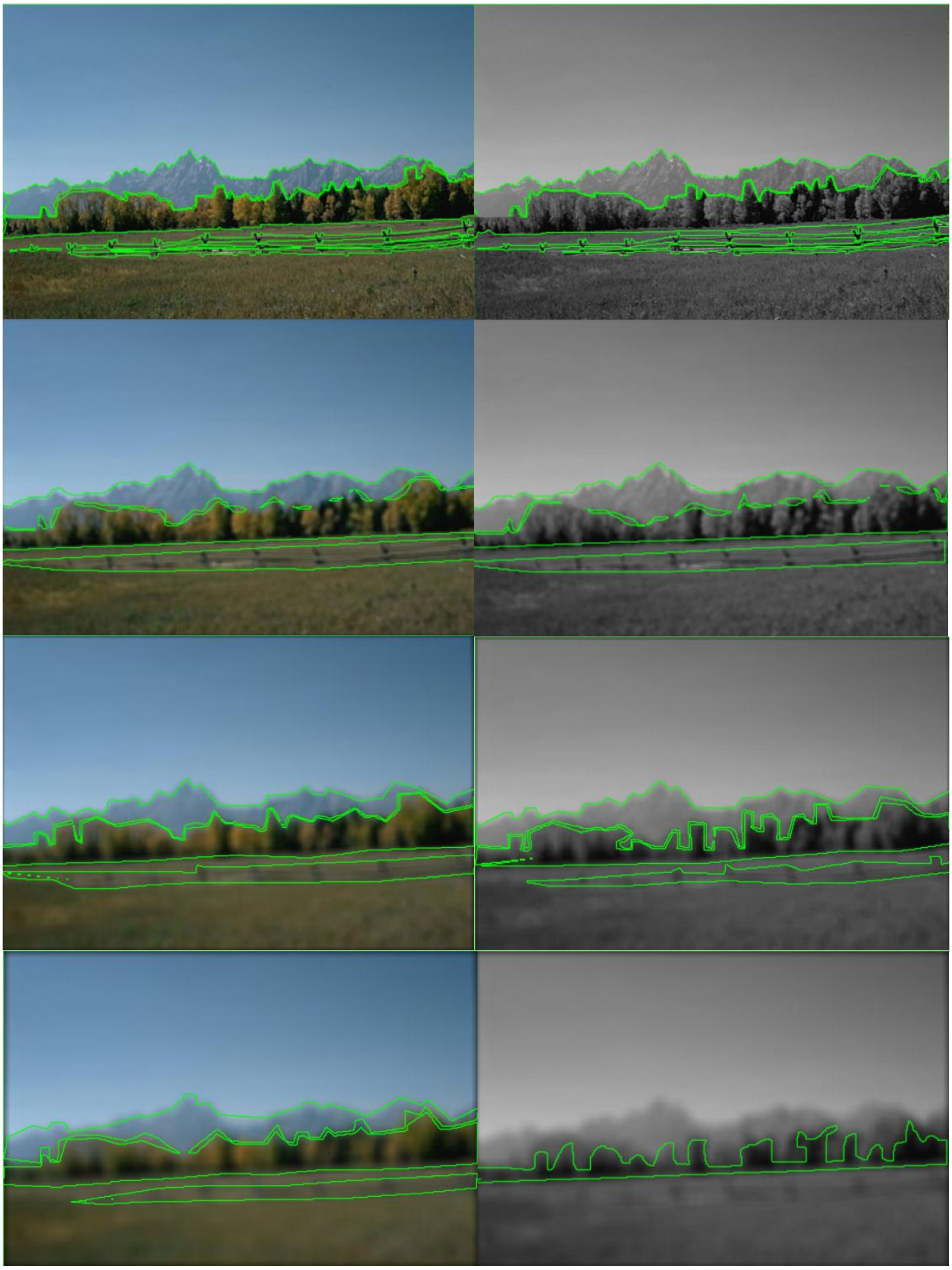


Figure 9. Example Human Segmentation for all four levels of blurriness (from zero to maximum) in colour and greyscale form

4.2. Human and Machine Segmentation Performance

In the process of testing hypothesis one, procedures of machine segmentation and evaluation were carried out. Approaches to each individual procedure were proved to be efficient.

As said before, human segmentation procedure successfully provided ground truth of human marked boundary maps, completing one of the main objectives of the project. This set of 800 in total segmentation files was checked upon collection, to ensure their validity and precision. Examples of human segmented images can be seen in Figures 9 and 16.

Having collected the human segmentation collection, experiments could move onto machine segmentation and from there to performance evaluation. Results of machine segmentation were gathered for every image in the system and for every level of blurriness, in both greyscale and colour.

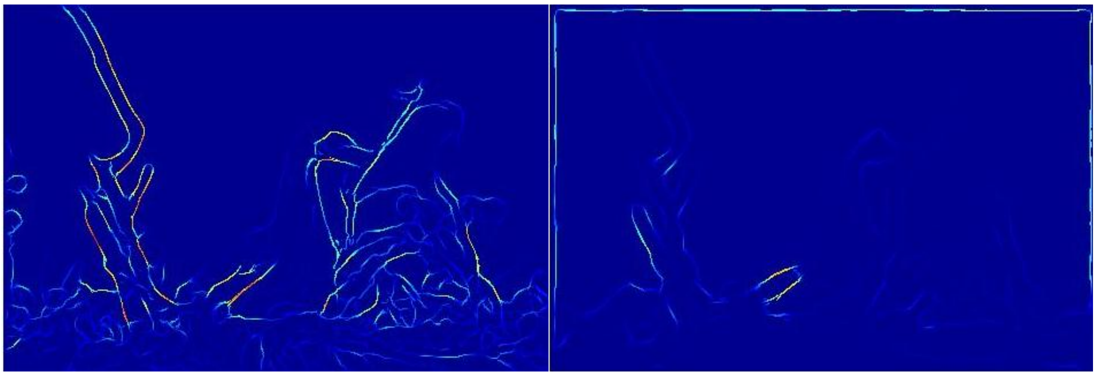


Figure 10. Machine segmentations on a coloured original image (left) and a coloured maximum blurred image (right)

Figure 10 displays two examples of machine segmentation. The left image in the figure illustrates machine segmentation of an image at its original format, while right image illustrates the machine segmentation of an image at maximum blurriness. Both of these machine segmentation examples are based on coloured images.

Results of machine segmentations were stored for every test image before being used to evaluate performance between human and machine.

As previously described, evaluation of human and machine performance was done using the framework of Precision Recall graphs. The experiment responsible for obtaining these outcomes, which combined machine segmentation and evaluation, was successfully completed. A Precision Recall graph was acquired for each level of blurriness for a particular image, different for colour and greyscale. These graphs were collected, stored and at the end compared to each other, both in regards of blurriness levels and colour/greyscale format as well. Figure 11 illustrates such a resulting graph, along with description of what each element of the graph represents.

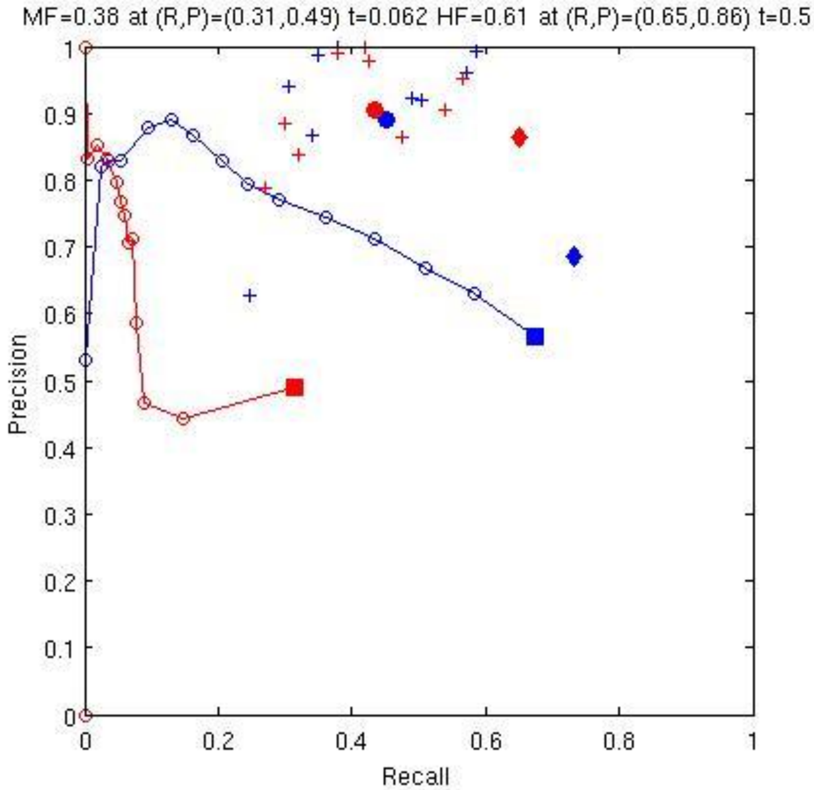


Figure 11. Combined Precision Recall Graphs: blue represents evaluation of performance on original image with zero blurriness, while red represents evaluation of performance when the image was blurred at maximum level. Dots represent mean value of human segmentations for each case, + represent human segmentations performance, o represent machine segmentation performance, squares represent the F-measure of machine segmentation and diamonds the F-measure for human segmentation.

An additional comparison present in the evaluation procedure was the mapping between machine and human segmentation. This was performed in order to visually compare the results from the two different segmentation techniques. Differences between human segmentations and machine segmentations were clear and precise. Results from mapping were visually displayed, separately illustrating the resulting machine segmentation and human segmentation, along with a match constituted from the points that both techniques had commonly provided. Figure 12 displays mappings of machine and human segmentations for one particular coloured image, for all four levels of blurriness.

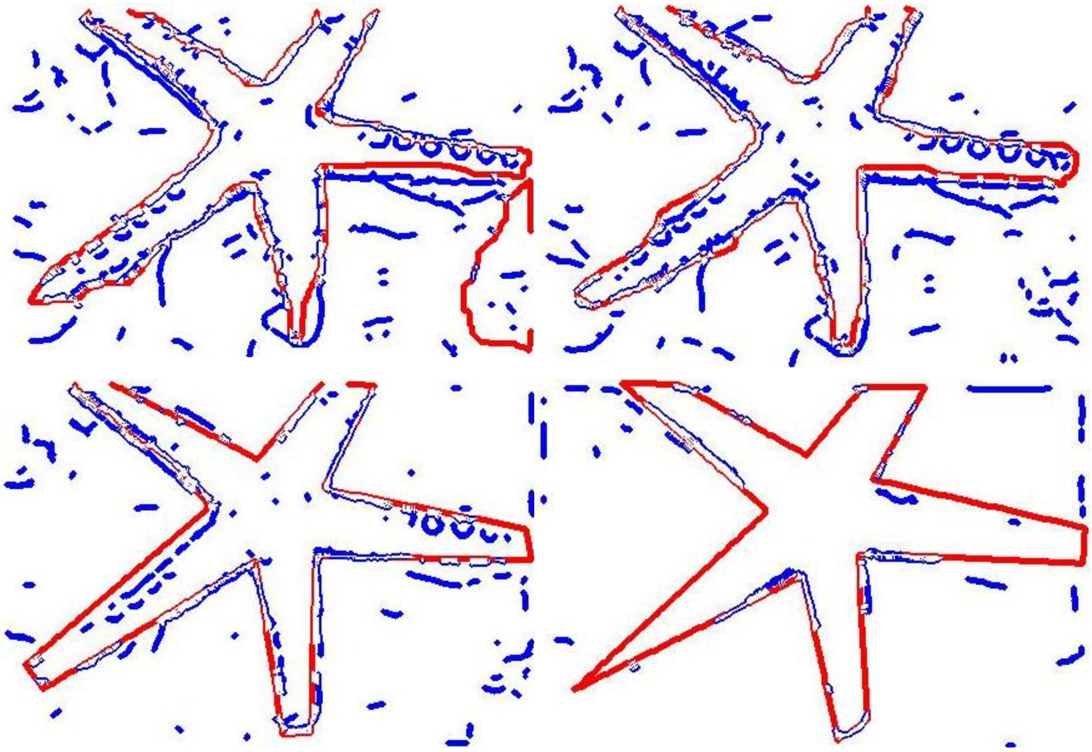


Figure 12. Matches between machine and human segmentations for all four levels of blurriness (moving row by row, left to right) Red denotes human segmentation marks, blue denotes machine boundaries and white lines show match between the two.

All of the above demonstrated experimental results were based on segmentations of coloured images. This is why the next section shall deal with the differences between results of segmenting colour images and grey images.

4.2.1. Greyscale Image Segmentation

Graphs in Figure 13 illustrate the PR graphs comparison between human and machine segmentation for two cases. Both graphs illustrate a comparison between original image and maximum blurred image machine segmentation. However, graph on the left demonstrates performance evaluation when image is coloured, while right image demonstrates performance evaluation when image is in greyscale. Both graphs represent evaluation of performance for the same image.

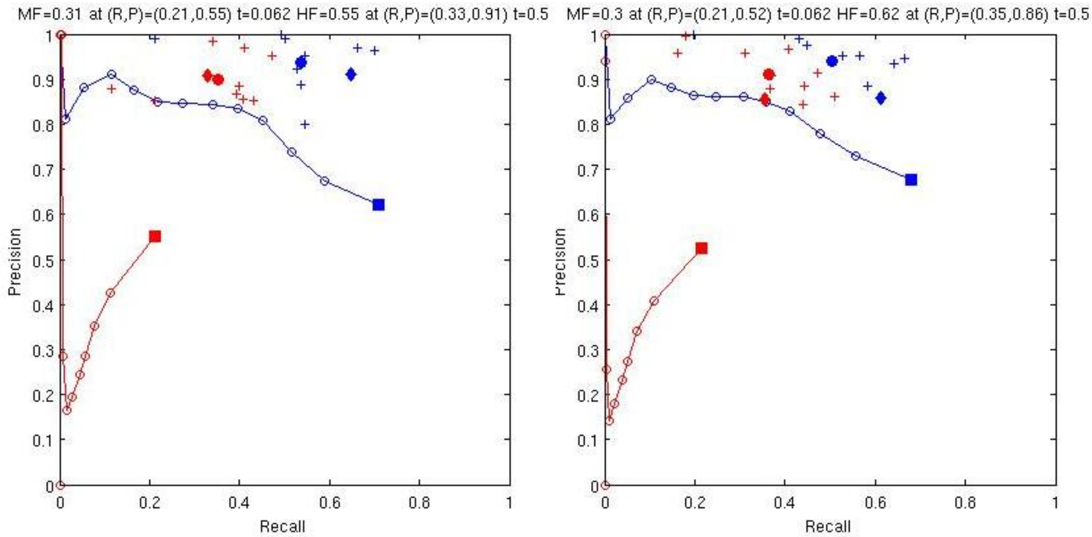


Figure 13. PR Graphs for original and blurred formats. Coloured(left) greyscale(right)

4.2.2. Aside

Before completing this section, it is worth having a look at the intermediate levels of blurriness and how this affected machine segmentation. The graphs in Figure 14 and 15 demonstrate machine performance when advancing the level of blurriness in the image. Graphs are dealing with coloured images.

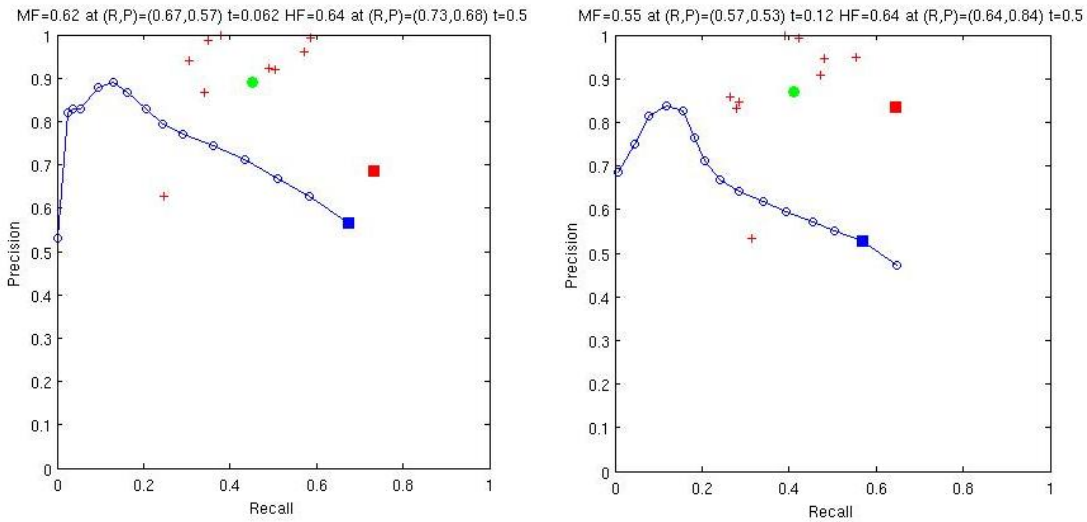


Figure 14. Original Image and minimum blurred image.

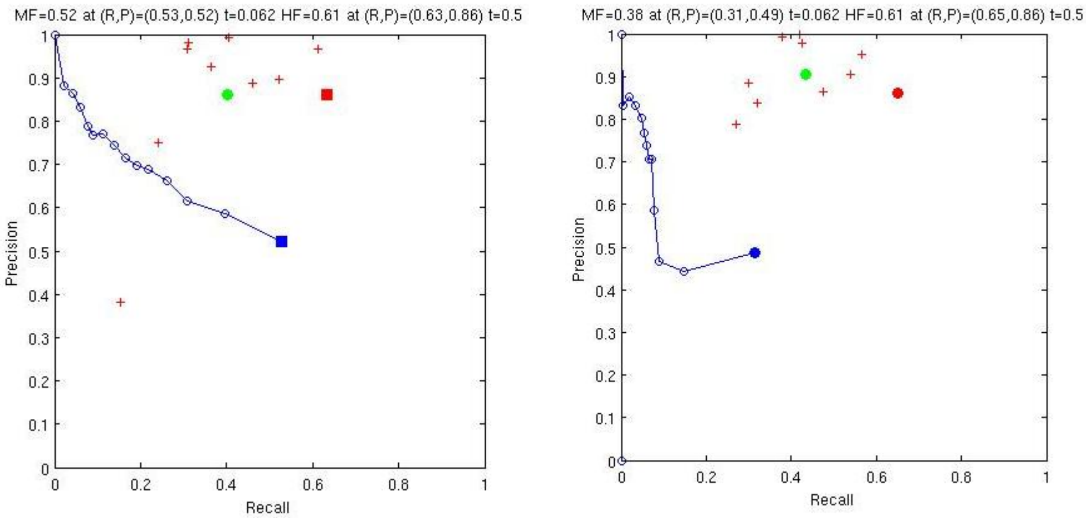


Figure 15. Medium bluriness image and maximum bluriness image.

4.3. Human Cognition of Images

Regarding the experiment of human cognition, the procedure chosen to produce the results was the most appropriate. Analysing observations from the resulting human segmentations, main points regarding human cognition were gathered. These were introduced into questionnaires, where they were completely supported by the participants.

It was observed that humans' cognition of an image is majorly affected by the complexity of the image, as well as its contents. Familiar objects, as well as images that consist of simple combinations of colours, shapes and textures seemed to be easier to identify at all levels of blurriness.

Analysis of segmentation of human participants with regards to image cognition was performed as designed. Image segmentations gathered were observed and compared to each other. Along with questionnaires that were distributed to the experiments' participants, data collected lead to results.

4.3.1 Observations

Through observation, it was clear that human segmentation varied at a great degree when moving from level to level in blurriness. As defined by the second hypothesis, the main concern in this experiment was to identify reasons for the variety in human segmentations. Aim was to identify characteristics of the image that would make someone perceive an image differently than another.

By comparing the original images segmentations with images that were blurred, (see Figure 16) for the whole set of images, certain characteristics causing particular difference in segmentation quality were identified. These characteristics were tested and confirmed to be the reasons for the cognition difference from the results gathered from the questionnaires given to each participant. Participants' responses are discussed in the next section, supporting particular features of images as parameters varying the quality of human image cognition.

Figure 16. Human segmentations on two images. The upper one is more complex than the lower one and this clearly affects segmentation of image at maximum blurriness.

4.3.2 Questionnaires Analysis

All participants completed the questionnaires, providing enough data to discuss relation of image content to segmentation. Table 1 presents the data gathered from the completion of the questionnaires, and more specifically, the amount of positive and negative answers.

It must be noted here, that questionnaires were completed by all participants. Each participant was asked to answer ten questions, of which the tenth included further comments on the experiment. For each question, participant had to verify (True) or deny (False) the given statement. No comments were provided by any participant regarding the segmentation procedures.

Table 1 Results gathered from Questionnaires

	Question	***True***	***False***
1	I found process of segmenting particularly easy.	9	1
2	I found segmenting original images easier than segmenting blurred images.	7	3
3	For each of my segmentations, I had no doubt that my results were as best as they could be.	2	8
4	I found segmentation of greyscale images more difficult than that of colour images.	6	4
5	I found segmenting images of familiar objects (e.g. seagull) to be easier than segmenting images of unknown objects.	8	2
6	I believe that complex objects, consisting of small components should be grouped together and not identified individually. (e.g. bird nest)	8	2
7	Images that were simple in colour combinations were easier to segment than more complex images.	9	1
8	Images that were simple in shapes combinations were easier to segment than more complex images.	10	0
9	Images that were simple in texture combinations were easier to segment than more complex images.	9	1

4.4. Summary

This completes presentation of experimental results. Overall, all experiments were successfully performed, providing all the necessary information to analyse the hypotheses. Next chapter analyses the obtained results, with particular focus to the validity of the hypotheses.

Chapter 5

Analysis of Results

Upon completion of results gathering, analysis on the information provided was performed. Analysis of results naturally leads to conclusions and to the final test of hypotheses validity.

5.1. Human and Machine Performance

Starting off with the most crucial topic analysed in terms of hypothesis one, discussion on difference of performance for human and machine segmentation shall take place. Graph in Figure 11 displays the difference in performance between human and machine segmentation, in both the case of a high resolution image and the case of a maximum blurred low resolution image. The difference in the two cases is clear. Machine segmentation performance has a significant decrease in the case of blurred image (red), while human segmentation does not change so rapidly.

The set of performances of human segmentations is distributed along a larger area for blurred images, while in original images the results are more closely together. The mean value does not change in a major degree, but still, it is higher in precision and lower in recall for blurred images. These minor changes show that working with blurred images does indeed affect human segmentation, but not in such a great degree. Humans still tend to agree with each other and provide similar results.

Also, F-measures for humans demonstrate a variance when moving from original to blurred image. As a reminder, F-measures show the trade off as a weighted harmonic mean of Precision and Recall. From graph in Figure 11, it is clear that F-measure of human segmentations moves into higher precision and lower recall for the blurred images. While the change for this value is more rapid than that of mean value, the results still have minimum change.

Completing discussion about effects on human segmentation, it can be said that human segmentation does indeed get affected by blurriness, but not in a high degree. Similarity between human segmentations seems less for blurred images rather than in the original images, but it is still high.

On the other hand, changes in machine segmentation are rapid when moving from original images to maximum blurriness images. The substantial effect that blurriness has on machine segmentation is clear in Figure 11. Performance of machine segmentation falls rapidly, resulting in both Precision and Recall being low.

Observing the results for machine segmentations on blurred images, it is easy to identify the reason for such low performance. In Figure 10, resulting machine segmentations on an original image and a maximum blurred image show that machine cannot properly identify boundaries of the image when this is in blurred format.

Machine segmentation on blurred images results in poor performance, as it seems that people are still able to identify the main objects in the image and therefore segment them. This is illustrated in Figure 12, once again in a comparison of machine segmentation and human segmentation results. Figure 12 illustrates the low performance of machine segmentation in comparison to human segmentation, which is visually obvious. When compared to human segmentations, machine segmentation cannot cope with the high performance of humans and therefore result into the low performance demonstrated in Figure 11.

As machine segmentations were collected and stored, differences between machine and human segmentation could be visually identified for all images. Significant difference between segmentations was clear at all the different levels of blurriness. The more blurred an image was, the less boundaries the boundary detector was able to recognise. However, humans were still able to comprehend objects in the image, and therefore successfully segment it. This resulted in the low precision match, where the little boundaries identified from the machine segmentor were matched to the human segmentation presented in Figure 12.

5.1.1 Greyscale Image Segmentation

Comparing results from colour images segmentation with that of greyscale images, it was discovered that there was no substantial difference between the two cases. Figure 13 supports this. Graphs in Figure 13 illustrate that greyscale in an image does not cause a large differentiation in the evaluation procedure. Therefore, conclusions regarding the first hypothesis can be drawn for images in general, without separating between colour and greyscale images.

5.1.2 Hypothesis Validity

From the graphs in Figures 14 and 15, it is clear that machine performance decreases as blurriness level increases. Difference between machine and human segmentation increases gradually, and finalises in the low performance demonstrated at the right graph of Figure 15.

What this change in machine segmentation performance implies, is that the first hypothesis is denied. Difference between machine and human segmentation increases as blurriness levels get higher, demonstrating a different behaviour than expected. While human segmentations do change and result in a more sporadic distribution, machine performance gets extremely low under the blurriness circumstances, so much that change in human segmentation performance is unnoticeable.

5.2. Human Cognition of Images

As shown in the results of human segmentations and the questionnaires, images that were simple in structure were segmented more successfully and easier than complex images. This supports second hypothesis of the project, showing that segmentation of images at blurred format is based on the cognition of the image. Cognition of the image in fact depends on its structure and content. The varying performances of humans in blurred images segmentations made it clear that images with different contents are comprehended in different manners.

These results open up a new notion in human cognition of images, introducing new approaches and observations. Approach followed towards this experiment was proved to be appropriate, resulting in observations that covered the whole area examined by the second hypothesis.

5.2.1 Analysis of questionnaires results

Analysing the answers participants gave in the questionnaires, belief that colour, shape and texture complexity in an image affected its segmentation is supported. Familiar objects were also more easily segmented than unknown objects.

Discussing the results more generally, we can start by supporting that segmenting an image is considered an easy process by most people. However, segmenting original images is also considered easier than segmenting blurred images.

What was worth noticing in the questionnaires was that most of the participants had doubts for their work. A striking 80% of the participants could not confirm that their work was as valid as possible. While most human segmentations are alike to each other, participants themselves did not consider their segmentations efficient.

As for difference between greyscale and colour images segmentation, the majority of participants believed that segmentation of colour images was easier than greyscale images.

When subjects were asked whether it was easier to segment images which had contents familiar to them rather than unknown, most of them (80%) answered positively. This provides an immediate connection of human cognition to how image segmentation is performed. It implies that objects familiar to participants were easier to understand and recognise, and it was therefore easier to segment them even at a maximum level of blurriness.

Another important question was the one related to understanding of human participants regarding the level of detail that should be reached at the segmentation of complex items. More specifically, participants were questioned about the case where objects constituted of smaller components. The aim was to check if the choice of segmenting a complex object as a whole instead of identifying every small component was conscious or done by mistake or ignorance. As it turned out, most of the participants (80%) believed that complex objects should be segmented as a whole instead of each component individually. Again, complexity is avoided and participants choose to segment in as simple manner as possible.

The last three questions were direct about three different image characteristics, colour, shape and texture. These three characteristics were identified through segmentations observation to control the complexity of an image, by relating their change ratio in the image, as well as their combinations. Indeed, the responses to the three last questions were amazingly supporting this notion. Almost all of the participants answered that simple colour and texture combinations made segmentation easier. Regarding texture, all of the participants, a significant 100%, believed that simple shape combinations were important to make segmentation easier. This implied that simple colour, shape, and texture structures in the image were better comprehended by humans and therefore made the segmentation process easier. Working with blurred images, making the major contents of an image even more significant, allowed us to observe that image cognition of humans seems to vary according to the colour, shape and texture complexity of the image.

5.2.2 Hypothesis Validity

From the procedure followed in the experiment and from the data gathered, a direct relation between the image complexities and image segmentation was clear. Consequently, image contents were also directly related to image segmentation.

5.3. Synopsis

Completing the whole of results analysis, it is possible to extract the main points. Starting with the first hypothesis, this was proved to be denied, as the difference between machine and human segmentation increased as the blurriness level became higher. Second hypothesis on the other hand was proved valid, as cognition of images is indeed affected by the contents of the image. Experiments chosen were appropriate to the purposes of the project, allowing for correct and complete gathering of information. Additionally, methodology applied was successful, providing high quality results.

Chapter 6

Conclusions

Naturally following the results' analysis, general conclusions of the project will now be stated.

6.1 Achievements

Overall, the project managed to successfully reach its goals, while following a continuous modification and proper adaptation as it progressed. All of the main topics, i.e. the main hypotheses, were thoroughly tested and analysed in order to verify their validity. As a reminder, the two hypotheses were briefly stated as:

Hypothesis 1:

"If blurred images are used in both machine and human segmentation, difference in their performance will be smaller than in the case of using high resolution images."

Hypothesis 2:

"If people are asked to segment a blurred image where detail is less, they will vary the quality of their segmentation from a high resolution image to the blurred one in relation to the contents of the image."

The most important achievement of the project is the provision of ground truth regarding human marked maps of blurred images. This has never been done before, a fact which makes this project the one to open up the way to further research. The dataset produced contains human segmentations on original high resolution images and three different levels of blurriness, both in colour and greyscale images. While a small dataset, it includes segmentations of ten participants, on ten different images for every level of blurriness, both in colour and greyscale. All of the information was collected, stored and is now available for further utilize. This is an excellent start for this section of image segmentation, regarding the time and resources limitations.

However, project was not based only on the human and machine segmentation data provision. The above dataset was used in the experimental machine segmentation on blurred images of the project, as stated in the first hypothesis. Executing machine segmentation on the blurred images and estimating segmentors performance with respect to the human segmentations on the blurred images, allowed the illustration of results regarding the association of human and machine segmentation when using blurred images. As thoroughly discussed in the chapter of analysis of results, the first defined hypothesis is rejected based on the data gathered. While the use of blurred images in image segmentation does indeed affect both machine and human segmentation performance, calculations of performance of machine segmentation show that machine segmentor performs badly in the case of blurred images. While there is an apparent change in human segmentation as well, machines performance decreases at an incredible degree. All of these lead to the conclusion that blurred images machine segmentation results in a poor performance by the currently implemented segmentors. This is always stated in comparison with performance in human segmentation. Until today, previous researches showed a performance of machine segmentors significantly lower in relation to human segmentation. Using blurred images, this performance difference was even more enlarged. The use of low resolution images, characterised by less amount of detail was proved to affect an image’s segmentation in both human and machine cases.

As far as the second hypothesis is concerned, this was proved to be accurate. By studying human segmentations, important differences in quality of segmentation related to the subject of the image were noted. This notion of analysis in image segmentation is also new, particularly focusing on the cognition of human subjects, not so much on a technical background. As part of the project, this section provided a new familiarity with the participants as individuals, providing a unique connection between the experiments investigator and the participants. Results covering this cognition analysis were not only based on the actual human segmentations but also on their various comments, reactions and beliefs during the process of segmentation. As previously described in chapter four of this document, the experiments followed the structure of [4], attempting to be vague and do not guide participants at any direction.

It is important to denote here that more than one participant asked the investigator if the way they were performing the segmentation was “correct”. This happened especially in the maximum blurred images segmentation, where all participants appeared to be uncertain of the validity of their work. These observations did not only show an amount of variety between general cognition of images, but also allowed for observation of cognition of images based on the image’s contents. Most of the participants were uncomfortable when dealing with a more complex image with different shades of colour, texture or shape and cases where many small sections made up a bigger one. The cases in which this happened were also important to note. As a general conclusion, it can be said that familiar and easily analysed complex objects were better segmented than unusual and difficult to comprehend objects. As a simple example, the segmentation of a human body was more detailed than that of a bird's nest.

Objects consisting of many extremely small components were consciously segmented as a whole instead of each individual component. On the contrast, segmentation of images with more explicit structure that did not contain large number of combinations of colours, textures or shapes were easily done in a more detailed manner.

It is unambiguous that project effectively and successfully covered all of the implicated notions. As an establishing project, it can represent the beginning of research into two previously unknown sections: image segmentation on blurred images and that of human cognition of images.

6.2 Critique

6.2.1. Overall

What can characterise the development of this project is a continuous spiral method, adapting objectives, system needs and resources from the first phase of analysis to the last gathering of results. As a new project, new ideas kept arising from both adapting previous work into this project's aims and interaction with experiment participants. This is why in comparing the initial objectives of the project with currents; a system evolved during development and changed into a combined technical and natural observation structure can be identified. Starting as a project aiming to expand on an existing research, repeat an existing approach to a new extend, it resulted in a project of new, additional analysis area and research.

Priorities of the system changed as well. As analysed in chapter two, design of appropriate user-friendly graphical user interfaces was given a great deal of attention. However, this was replaced with more technical hypotheses to test the data against. More specifically, in the initial specifications of the project, analysis regarding human cognition of different objects was not included. This feature of the project was added later on, inspired by the resulting machine segmentation performance and user experiments.

While this nonstop change resulted in overhead during the project progress and sometimes lose of time, the resulting system and project are of excellent quality. Nevertheless, based on the nature of the project as a research based one, adaptability was not an unexpected requirement. Despite the fact that the essential objectives related to system coherence and easier use by users were not covered, this is acceptable as this is a project specialised in a computer vision area, and was mainly technical. Therefore, overall it can be said that project covered all of its objectives satisfactorily, extended efficiently in the required sections and prepared an excellent basis for further research.

6.2.2. Design of Experiments

This section includes changes applied to experiment design and approach.

To begin, discussion on how the final approach for the human segmentation experiment was carried out is introduced. Approaches to this experiment changed, as while experiments were carried out, room for improvement was discovered.

Human Segmentation Experiment

Human segmentation experiment was approached in a different manner, as the original design was proved to be extremely time demanding. Through discussion, both investigator and participants decided to apply different techniques. The following sections include the analysis of the various approaches applied before deciding on the final one.

- **Human Segmentation Experiments Approach II**

As the initial approach was proved impractical and inefficient, it was decided to organise lab experiments, where each participant could work individually, in a friendly environment while still get as much as possible support and supervision. This solution was intended to save time both to the investigator and the participants. However, technical difficulties did not allow for organisation of lab sessions, therefore a totally different approach was introduced. For this final approach, participants were contacted and asked about what they thought as a best approach.

- **Human Segmentation Experiments Final Approach**

After discussing with the participants for what approach would suit them first, the most suitable way to perform the experiments was applied.

For each participant, the java application for human segmentation was installed on his/her personal computer. Due to the familiar and light structure of the system, this was an easy procedure that only required one visit per participant. Setting up the application did not take much time. Of course set up was completed by the investigator on the participant’s personal machine. Another characteristic of the java application was its adaptability on operating systems. After setting up the system, this procedure was never to be repeated.

In the first meeting with the participants, investigator gave a small presentation on how to use the system. After that, participants performed some example of the needed procedure, just for the investigator to check that they understood the use of the system completely. This was completed by allowing the participants to experiment with the software for a short period of time, in order for them to try to perform all the actions required to run the experiment, which were also presented in the investigator presentation. Any misunderstandings or questions could be raised at this point, ensuring the participants understanding of the project.

From there, each participant was free to work whenever and for as long as he/she wanted. No continuous communication or physical transfer was required from any of the two sides of investigator and participant. While each participant was left to perform the experiments alone, supervision and support were still considered important in the project.

The fact that participants had different backgrounds, with some of them having less experience in interaction with programs, made the topic of support availability important. Moreover, as in the duration of the experiments a holiday period was present, other ways for interaction, other than physical communication were required. Difficulties could arise at any point, as participants could forget the procedure or accidentally change something, resulting to a faulty behaviour of the system. Therefore, the investigator had to be available and reachable any time of the day and under any circumstances. This was easily dealt with the use of digital media communication. Emails, video calls, instant messaging, social networking sites and telephone calls were the tools used for distant communication. Participants could use any of those tools to contact the investigator, at any time. Immediate responses were possible almost for every case. Most issues were resolved through the distant communication, but whenever necessary, an appointment was arranged for the investigator to check the problem first hand. Through this guidance and support availability, human segmentation experiment was successfully completed.

Despite all the changes in experiments' structure, the final approach proved to be both the most efficient and most time saving. Of course it had some pitfalls, for example non complete supervision of participants, asynchronous support and distance communication. On the other hand, participants were able to ask for support any time, through instant messaging tools, telephone calls, video conferencing and asynchronous messaging. Therefore, contact with the participants was not completely lost. Additionally, while the investigator was not present during the whole of experiments, this did not eliminate the opportunity to gather information and comments from the participants. This was not only achieved through the constant support, but also via the physically supervised introduction to the system performed by the investigator to every single participant. The first segmentations were done in the presence of investigator, a test to check whether participants fully understood the system and its functionalities. Furthermore, this allowed for an interaction between participant and investigator regarding any questions, comments and general thoughts.

The supervision method adopted in this project cannot be characterised neither as better than physical supervision, nor worse. It was of course very different for both the investigator and the participants. Through discussion of these issues with Dr L. Watts, it was decided that in order to get a complete critique of this approach, both potential advantages and disadvantages shall be discussed.

Starting off with advantages, communication via computers can be less intimidating for people. This means that participants could be more confident when dealing with the investigator. They would not be shy or nervous to ask for help, something which could be the case in physical communication. Moreover, both participants and investigator had more flexibility in arranging when to communicate. This means that participants using asynchronous messaging would be able to reconsider their question and contact investigator whenever they considered best. Additionally, as in some cases investigator was in different time zone than participants, no synchronous communication was possible. Participants could easily form their questions and investigator would attend them as soon as possible, or reply to arrange when to talk. Another advantage in using distant supervision was that conversation was possible to be recorded (IM log or social networking sites messages/emails). This enabled the participants to review the discussion later, in case they forget or misunderstand the explanation at the time of discussion.

Moving on to disadvantages, we can start with the fact that emotional communication through computers is of course harder for people that do not know each other very well. This feature did not appear to affect the project so much, as a closer relationship was gained with each participant before moving on to distant supervision. Therefore, this disadvantage does not fully apply in this case. Another disadvantage of distant supervision is that while explaining about physical things, i.e. the software user interface, it is not easy to identify things, as gestures are not feasible. However, some of the means used to support distant supervision provided a solution to this difficulty. An example of that is video call's functionality of sharing the screen of your computer. Using this functionality, it was easier for the investigator to guide participant. An example where this feature was useful is the case of dealing with software issues on the participant's system installation on his/her personal PC. Using screen sharing, investigator was able to visually guide the participant to perform the appropriate actions necessary to resolve the issue.

Nevertheless, there were cases in which video call was not enough to identify the problem. In such cases, a physical meeting between investigator and participant was arranged for the investigator to personally resolve the issue causing the problem. A last disadvantage of distant support is the fact that recorded conversations could be used by participants to quote back to someone out of context parts of the conversation, e.g. rude statements. If something like this happened in the selected approach of distant supervision, investigator would probably be unaware of this forever, unless such an event caused an issue.

As it can be drawn from the above discussion, most disadvantages of distant supervision were approached in some degree through the various media used for communication. This means that while this method still implies some disadvantages and difficulties, it can be considered efficient when these difficulties are overcome and no physical supervision is necessary.

Despite the distant communication, participants and investigator developed a friendly relationship through support and attendance to participants' problems. From the side of view of the investigator, it was completely understandable that participants also had other obligations and were therefore not able to focus entirely on the experiment. However, participants' interest in the project increased with their participation of the program, and most of them asked how they could view their work, how exactly it would be used and asked to see the final results.

Machine Segmentation and Evaluation Experiments

Initial designs of the experiment combined machine segmentation with segmentation evaluation procedure. This was indeed the final approach to the experiments, but an intermediate level of approaches that was applied, tested and finally dropped shall now be discussed.

In the stage of acquiring the machine segmentation results, a new method was discovered, again from the same research team in [7]. This included code for running BSE on its own, providing files with information about machine segmentation. BSE did not have any particular requirements, running on Linux without any additional software required. In the initial approach to this experiment, use of BSDS demanded Matlab environment installed on Linux. As Matlab is not a free software package, and an alternate way was available, using BSE directly seem ideal. Additionally, BSE's type of resulting machine segmentations seemed to be identical to the results of human segmentation from the system manual. This meant that analysis procedure would be easier and more efficient.

However, when BSE was applied to gather the machine segmentation results, it was proved that while the resulting file types had the same file extension (.seg) as the human segmentation files, the contents and format of the files were different. This made the evaluation of both types of segmentation complex, making use of BSE a bad choice for the purposes of the project.

Proved to be wrong, the initial design of the experiments was reinforced and followed. It was indeed proved to be the most ideal approach, as BSDS provided all the functionalities needed and was simple in understanding and using. Machine segmentation and evaluation procedure was finally completed successfully.

Overall, experiments were the most pleasant part of the project development. Successfully, the best approaches to minimise time consuming were used, while, changing between approaches to experiments did cause some time lose.

6.2.3. System Development Process

As expected, process of understanding, using and combining all of the available software required more effort and time than estimated. Three different codes were used in the contents of the project: the java application supporting human segmentation, the BSE machine segmentation [5] and the Berkeley Segmentation Dataset and Benchmark (BSDS) [7].

A number of difficulties came up when working with the software. First of all, setup of some of the applications took more time than expected. This mainly applies to the BSDS, as not only it required specific environment to be able to be installed in, but several software and applications also. Additionally, most of the functions provided in BSE are included in BSDS Benchmark. BSE was used to provide machine segmentations, but was then replaced by using BSDS' functionalities combining machine segmentation and analysis.

Another important fact related to the system development was the use of different formats and structures and operations of these different software packages. Java application for human segmentation outputted the results into a file of type .seg, defined as a text file including data analysing every part of the image line by line and the segment to which it was assigned. On the other hand, BSE also outputted .seg files, which had totally different format than the java application. In this case, .seg files were in reality a matrix with dimensions equal to those of the image and element's values the segments numbers to which each element belonged. Changing between BSDS to BSE and back to BSDS cost the project valuable time.

An additional issue regarding the code was its age. Using a relatively old (while still significant) project, dated in 2004 [6] as guideline and using the same software packages related to it, was of course related to some pitfalls. After dealing with some issues related to the code, particularly Java application, the implementers of the code and guideline project [6] were contacted. From there a new, updated version of the java application dealing with some errors was accessed, a source unavailable on the Internet. Another situation in which the age of the guideline document was obvious was in requests for support from the implementers of the used software in regard with some issues, and getting replies describing how long it has been since they last saw the code and how they would need time to refresh their memory.

Concluding, the fact that a variety of media and software environments were used for the development of the project raised various matters. These made the analysis complicated and required time to adjust data form between the different applications. However, in the resulting project, everything was completely resolved and defined. The only side effect this issue caused was the time waste when moving from one approach to another.

6.2.4. Research Process

Examining the research process followed, it can be said that everything ran smoothly concerning the theoretical part. The paper of [6] was proved to be an excellent guideline, providing all the necessary information and resources for better understanding. The part of the project that required the most research was that of Precision-Recall graphs. This is natural, as these graphs define the evaluation methodology used in the project and should therefore be analysed as much as possible.

As far as the execution of the experiments is concerned, the original source of the java application for human segmentation [4] provided outstanding guidelines for user approach and experiment’s structure. These were proved to result in excellent quality experiments, giving the appropriate importance to leave the participants unaffected by supervision and follow their own personal cognition through the whole of the segmentation process. It was therefore confirmed that the choice of the particular resource for experiments guidelines could not be better.

As the practical part in research is concerned and since the total of code used in experiments was open source code, it was important to understand how everything worked. Combining it with theory, everything should become clear. Of course, this was not completely established, as in most of the cases, in order to understand a theory, you also have to apply it. However, it was not difficult to comprehend the structure of all three available software packages and successfully use them for the purposes of the project.

6.3 Personal Learning

Working on a research project of a topic that was unfamiliar before, was at least a challenging task. Upon completion of the project, it is a delight to see how much knowledge and understanding it offered. This project provided a great deal of personal gain in knowledge that would not be gathered in any other way. First of all, this project was never done before. This allowed for the preparation of data and observations new in the area, not only for individual understanding but for other researchers to use as well.

Probably the most important feature of developing this project was that the implication with research in computer vision. Before this project, nothing was known regarding previous and current research in image segmentation. Results and conclusions related to machine and human segmentation association was also unknown. Now, most of the important work done in image segmentation is familiar. Experience in working with the currently best segmentor in the world and understanding the mathematics and procedures behind it are also two really important skills gained from the project. The rest of image segmentors, the algorithms behind them and their final performance is also well-known. Last but not least, the use and construction of Precision-Recall curves was perfected. This is an exceptionally important skill, as this method of evaluation of performance is one of the most popular today.

As this project explored a new section of image segmentation, trying to extend available research, its results and conclusions are pioneer and can constitute the beginning of a greater research. Using previous work of one of the most significant research teams in the world and discussing with them, provided a good networking environment. Now, more than ever, it is clear that in developing a project, implementers will always be "standing on the shoulders of giants". It is very hard to perform individually. Talking to fellow researchers, open to new approaches, advices and suggestions should be part of any research. In contacting other researchers for this project, everybody was interested and willing to provide help and support. This included guiding to other people for specialised in the requested area. Through this, new people implicated in the continuation of previous research were met.

Moving on to the non technical gains, experiments organisation and supervising is also considered a valuable skill from this project. Working with many people always practises communication skills, user approach and allows for new ideas through the users behaviour. Ethics related to experiments were an important notion in the project too. Studies regarding ethics matters of "use" of people and their application provided essential knowledge and understanding of the ethical implications in performing experiments.

6.4 Redesign

In a proposed redesign of the system, several things could have been done differently.

Firstly, before deciding the final approach to experiments, discussion with the participants should take place. This would allow for acknowledgment of any difficulties and recommendations. The most suitable approach would be drawn from such discussion, enabling the optimum time management of experiments and pleasant and more fitting execution of the experiments for both the investigator and the participants. Moreover, the group of participants selected to perform the experiments would be chosen from a team of people with lighter schedule, having the necessary time and strength to provide results of top quality. In this project, while people performed sufficiently well in the human segmentation, it was obvious that their busy schedule affected both their effort and enjoyment of the process. Under better circumstances, participants would be able to dedicate to the experiments, actively participating.

Regarding the more technical aspects of the project, especially the software coordination in the project demands, approach would also be differentiated in a redesign of the system. In such case, better communication with people using the same technology would be targeted, talking about the project aims and gathering suggestions, solutions. This would help to gain a better understanding of the problem and be informed of new techniques available. Additionally, better management of resources would also be achieved.

Finally, in a complete redesign of the system, a platform would be implemented in order to combine all the necessary features into one application. This would enhance the users to interchange between the available functionalities and get the best analysis.

6.5 Contributions to Current Knowledge or Practice

This is a pioneer project in its notion, establishing the beginning of further research. Both of its hypotheses included new ideas in the area of computer vision, notions that were never dealt with before. The findings of this project are significantly important, providing a basis for future research.

Additionally, the dataset of human marked maps gathered from the experiments defines ground truth. This would be extremely useful in further research, allowing other researchers to use it to base their analysis on them.

6.6 Possible Future Work

An initial extend on this project would be the testing of the results on a larger scale. This shall include more participants, volunteers from every age and background so that no limitations would appear in the analysis framework.

Additionally, all available boundary detectors should be tested on blurred images segmentation. Their performance could change from segmentation on natural images to segmentation of blurred images. Order on which they appear from the most efficient to the least efficient could also change.

The currently available dataset should be uploaded to the Internet, where other researchers would be able to access it for free and use it for further research. Something like that would follow the structure of the BSDS [7], only this time implementing a dataset of blurred images rather than images of high resolution.

Another possible extent to this project would be the use of several different measures of segmentation similarity. Based on [29], a recent paper discussing and comparing several different measures of segmentation similarity as well as different segmentation datasets, this project could be included in the same framework. More techniques in measure of segmentation similarity could be approached and tested against blurred images.

Another feature that could extend the current project would be more specifically focused on the cognition of blurred images. This involves extending the second hypothesis analysis into more detailed discussion, checking which image features were responsible for the cognition observed.

Finally, another important factor that came up in the duration of the experiments and from the various observations, could be to test whether the sequence in which participants performed experiments affected their cognition of the image. In the current project, it seemed that when participants started segmenting from the original image and then moved on to the most blurred one until they reach the maximum blurred one, their segmentation was not as truthful and trustable. This is believed to be because participants "knew" what was in the blurred image from segmenting all the better qualities of it. Participant following the above order in segmenting the experiments would segment the images in a way they probably would not under normal circumstances. This was also observed during the experiments, where some participants commented that despite the fact that they could not properly identify an object in the image, they considered it correct to segment it because they "knew what was there" from the previous segmentations.

6.7 Summary

Overall, this pioneer project was defined by excellent quality and results. Following appropriate guidelines from previous works, it used available software and produced results and observations new for the area of image segmentation. Being a basis for further research, this project could be used to support other researchers. The ground truth it provides, along with the analysis and opportunities for further research make it an ideal starting point for many differently focused potential future projects.

Bibliography

[1] Puranik, P. et al., “Human Perception-based Color Image Segmentation Using Comprehensive Learning Particle Swarm Optimization,” *Second International Conference on Emerging Trends in Engineering and Technology*, ICETET-09, ISBN 978-0-7695-3884-6, IEEE Press, pp. 630-635, 2009

[2] Vazquez, E., Baldrich, R., Vazquez, J., Vanrell, M., “Topological Histogram Reduction towards Colour Segmentation Source,” Lecture Notes in Computer Science; Vol. 4477, *Proceedings of the 3rd Iberian conference on Pattern Recognition and Image Analysis*,Part I

[3] Mangin, J.-F., Tupin, F., Frouin, V., Bloch, I., Rougetet, R., Régis, J. and Lopez-Krahe, J., “Deformable topological models for segmentation of 3D medical Images”, *14th Information Processing in Med. Imaging (IPMI),* Brest, France, Computational Imaging and Vision, pages 153--164, Kluwer Academic Publishers, 1995

[4] Martin, D., Fowlkes, C., Tal, D., and Malik, J., “A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics,” *Proc. Int’l Conf. Computer Vision*, 2001

[5] Martin, D., Fowlkes, C., and Malik, J., “Berkeley Segmentation Engine,” available from Internet [http://www.cs.berkeley.edu/~fowlkes/BSE/]

[6] Martin, D., Fowlkes, C., and Malik, J., “Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues”, *IEEE Transactions On Pattern Analysis And Machine Intelligence*, Vol. 26, No. 1, pp. 530-549, January 2004

[7] “Berkeley Segmentation and Boundary Detection Benchmark and Dataset,” 2003, available from Internet [http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/]

[8] Schalkoff, R., Shaaban, K., and Carver, A., “Image Processing Algorithm Design and Implementation for Real-Time Autonomous Inspection of Mixed Waste”, Department of Electrical and Computer Engineering, Clemson University, 1996

[9] Fowlkes, C., and Malik, J., “How Much Does Globalization Help Segmentation?”, *Report No. UCB/CSD-4-1340*, Computer Science Division (EECS), University of California Berkeley, July 2004

[10] Morrone, C., and Burr, D., “Feature Detection in Human Vision: A Phase-Dependent Energy Model”, *Proc. R. Soc. Lond. B 235*, pp. 221-245, December 1988

[11] Mallows, C., “A Note on Asymptotic Joint Normality,” *Annals of Math. Statistics, vol. 43, no. 2*, pp. 508-515, 1972

[12] Levina, E., and Bickel, P., “The Earth Mover’s Distance is the Mallows Distance: Some Insights from Statistics,” *Proc. Int’l Conf. Computer Vision, vol. 2*, pp. 251-256, 2001

[13] Palmer, S., Vision Science: Photons to Phenomenology, MIT Press, Cambridge, 1999

[14] Malik, J., Belongie, S., Leung, T., and Shi, J., "Contour and Texture Analysis for Image Segmentation," *Int'l J. Computer Vision, vol. 43, no. 1*, pp. 7-27, June 2001

[15] Puzicha, J., Hofmann, T., and Buhmann, J., "Non-Parametric Similarity Measures for Unsupervised Texture Segmentation and Image Retrieval," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 267-272, 1997

[16] Julesz, B., "Textons, the elements of texture perception, and their interactions", Nature Publishing Group, pp. 91 – 97, doi: 10.1038/290091a0, March 1981

[17] Van Rijsbergen, C., Information Retrieval, second ed., Dept. of Computer Science, University of Glasgow, 1979

[18] Davis, J., Goadrich, M., "The Relationship between Precision-Recall and ROC Curves", Department of Computer Sciences and Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison

[19] Martin, D., Fowlkes, C., and Malik, J., "Learning to Detect Natural Image Boundaries Using Brightness and Texture," *Neural Information Processing Systems*, 2002

[20] Levina, E., "Statistical Issues in Texture Analysis," *PhD thesis*, University of California, Berkeley, 2002

[21] Schapire, R.E., and Singer, Y., "Improved Boosting Algorithms Using Confidence-Rated Predictions," *Machine Learning, vol. 37, no. 3*, pp. 297-336, 1999

[22] Jordan, M.I., and Jacobs, R.A., "Hierarchical Mixtures of Experts and the EM Algorithm," *Neural Computation, vol. 6*, pp. 181-214, 1994.

[23] Chang, C., and Lin, C., LIBSVM: A Library for Support Vector Machines, 2001, available from Internet [http://www.csie.ntu.edu.tw/cjlin/libsvm]

[24] Oppermann, R., "User-interface design", Institute for Applied Information Technology, GMD Forschungszentrum Informationstechnik

[25] Nielsen, J., Usability Principles, 2001

[26] Haubold, A., *Aquaphoenix lectures*, University Of Columbia, available from Internet [http://www.aquaphoenix.com/lecture/matlab10/page3.html]

[27] The MathWorks, Inc., MATLAB® Graphics, *Online only Revised for MATLAB 7.9 (Release 2009b)*, September 2009

[28] The MathWorks, Inc., MATLAB Creating Graphical User Interfaces, *Online only Revised for MATLAB 7.0 (Release 14)*, June 2004

[29] Arbelaez, P., Maire, M., Fowlkes, C., Malik, J., "Contour Detection and Hierarchical Image Segmentation", Electrical Engineering and Computer Sciences, University of California at Berkeley, Technical Report No. UCB/EECS-2010-17, February 16, 2010, Available from Internet [http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-17.html]

[30] University of Bath, Department's of Computer Science Research Ethics checklist, available from Internet [http://www.cs.bath.ac.uk/Leon/pages/ethics.shtml#checklist]

[31] Watts, L., Research Ethics, Lecture for CM30082 unit, 2010

[32] Mitscherlich, A, Mielke, F., The Nuremberg Code (1947) In *Doctors of infamy: the story of the Nazi medical crimes*. New York: Schuman, 1949: xxiii-xxv

Appendix A

Research Ethics

UNIVERSITY OF BATH

Department of Computer Science

13-POINT ETHICS CHECK LIST

This section describes thirteen issues that were carefully considered before involving any other people ("participants") in this project for the purposes of collection of information.

1. *Have you prepared a briefing script for volunteers?*

In order to explain to participants what they will be required to do, what data will be collected from them and how these data shall be used in the project, an online briefing script shall be provided. Briefing script will be visible online so that any participant can access it at any time during the experiments. Script includes information on how users should store the data collected and how to organise their results for their usability benefits. Additionally, explanation on how the data collected shall be used will be was visually presented to the participants in an introductory presentation of the system and its use, performed to each of the participants individually.

2. *Will the participants be using any non-standard hardware?*

Participants will not be exposed to any risks associated with the use of non-standard hardware. Typical interaction with their personal laptop or occasionally another machine will be required for completing the experiments.

3. *Is there any intentional deception of the participants?*

No intentional deception of the participants shall be present.

4. *How will participants voluntarily give consent?*

Participants will be asked to sign a consent form regarding the use of the results of their experiments beyond the completion of this project. This is because data gathered from the experiments shall be published in the future. A separate consent form will be signed by each participant.

5. *Will the participants be exposed to any risks greater than those encountered in their normal work life?*

Project only contains a minimal amount of risk, which is highly unlikely to take place. As participants will be working on low resolution images, this could tire their eyes. However, these experiments do not involve any danger though, as it is performed only for a limited amount of time. Professional opinion was advised, in order to confirm all of the above.

6. *Are you offering any incentive to the participants?*

No harm risk shall be implicated to the experiments. Participants shall only be rewarded with the organisation of an event in order to thank them for their participation.

7. *Are any of your participants under the age of 16?*

None of the participants is under the age of 16.

8. *Do any of your participants have an impairment that will limit their understanding or communication?*

None of the participants has any impairment that could limit their understanding or communication.

9. *Are you in a position of authority or influence over any of your participants?*

No. No pressure shall be present to the participants to take part in, or remain in any experiment.

10. *Will the participants be informed that they could withdraw at any time?*

All participants will have the right to withdraw at any time during the investigation. This shall be included in the briefing script.

11. *Will the participants be informed of your contact details?*

All participants will be able to contact the investigator anytime during and after the investigation. Both asynchronous and synchronous communication will be enforced. Participants will be able to contact the investigator any time using a variety of different digital mediums, including instant messaging and social networking sites. All of the participants shall also be able to contact my supervisor via e-mail or telephone.

12. *Will participants be de-briefed?*

Participants will be informed of the use of data collected from them. They will also have the opportunity to ask any questions.

13. *Will the data collected from the participants be stored in an anonymous form?*

Personal data of the participants will not be passed on, but only be used for relevant purpose.

Association of Human and Machine Segmentation Using Blurred Images

NAME: Maria Lena Demetriou

SUPERVISOR (IF APPLICABLE): Peter Hall

SECOND READER (IF APPLICABLE): ______________________________

PROJECT TITLE: Association of Human and Machine Segmentation Using Blurred Images

DATE: ____

Appendix B

Code

Blurring Code

MyBlurring.m

```
close all;
clear all;
clc;
count = 0;
while count<10
    %% Setting Blurring Properties
    infile = num2str(count);
    filename = ['Original Images/', infile, '.jpg'];
    sigma = 8;
    hsize = 5;
    %% Reading Image in
    im1 = double(imread( filename ))/255;

    i=0;
    while i<3
        %% h = fspecial('gaussian', hsize, sigma)
        h = fspecial('gaussian',hsize,sigma);
        im2 = imfilter(im1,h);
        %% Displaying figures
        figure;
        imshow( [im1 im2] );
        %% Saving Blurred Image
        imwrite(im2, ['BlurredImages', num2str(hsize),'/', infile,
'Blurred.jpg']);
        hsize = hsize+5;
        i=i+1;
    end
    count=count+1;
end
```

Machine Segmentation and Evaluation

graphMain.m

```
% Control what to test from here
% The main control of the program

%Preparing environment
clc;
clear all;
close all;

[MR,MP,mf,mthresh,HR,HP,hf,hthresh] = analysis(30000,'gray');
[bMR,bMP,bmf,bmthresh,bHR,bHP,bhf,bhthresh] =
analysis(300015,'gray');

figure(10); clf;
plot(MR,MP,'-bo');
hold on;
plot(HR,HP,'b+');
% Ploting human segmentations' mean
plot(mean(HR),mean(HP),'bo','MarkerFaceColor',[0 0
1],'MarkerSize',7);

plot(bMR,bMP,'-ro');
plot(bHR,bHP,'r+');
% Ploting human segmentations' mean
plot(mean(bHR),mean(bHP),'ro','MarkerFaceColor',[1 0
0],'MarkerSize',7);

axis equal; axis([0 1 0 1]);
xlabel('Recall'); ylabel('Precision');

% find best MF and HF-measure (should interpolate)
[t,idx] = max(mf(:));
[ht,hidx] = max(hf(:));
title(sprintf('MF=%.2g at (R,P)=(%.2g,%.2g) t=%.2g HF=%.2g at
(R,P)=(%.2g,%.2g) t=%.2g',...
              mf(idx),MR(idx),MP(idx),mthresh(idx),
hf(hidx),HR(hidx),HP(hidx),hthresh(hidx)));
plot(MR(idx),MP(idx),'bs','MarkerFaceColor',[0 0 1],'MarkerSize',7);
plot(HR(hidx),HP(hidx),'bd','MarkerFaceColor',[0 0
1],'MarkerSize',7);

[t,idx] = max(bmf(:));
[ht,hidx] = max(bhf(:));
title(sprintf('MF=%.2g at (R,P)=(%.2g,%.2g) t=%.2g HF=%.2g at
(R,P)=(%.2g,%.2g) t=%.2g',...
              bmf(idx),bMR(idx),bMP(idx),bmthresh(idx),
bhf(hidx),bHR(hidx),bHP(hidx),bhthresh(hidx)));
plot(bMR(idx),bMP(idx),'rs','MarkerFaceColor',[1 0
0],'MarkerSize',7);
```

```
plot(bHR(hidx),bHP(hidx),'rd','MarkerFaceColor',[1 0
0],'MarkerSize',7);
hold off;
```

analysis.m

```
function [MR,MP,mf,mthresh,HR,HP,hf,hthresh] = analysis(iid,present)

% User inserting image id and type (colour/gray)
% iid = input('Enter the id of the image to be analysed: ');
% present = input('Enter "color" or "gray" depending on image type:
', 's');

% setup
userno=10;
HP = [];
HR = [];

% Calculating machine's segmentation PR values
disp('Calculating machine segmentation...');
[MP,MR,mf,mthresh] = machinePR(present,iid);
disp('done.');

% Calculating all human's segmentation PR values
disp('Calculating all human segmentations...');
for i=1:userno
    uid = i+1000;
    [hp,hr,hf,hthresh] = humanPR(present,iid,uid);
    HP = [HP hp];
    HR = [HR hr];
end
disp('done.');

figure(); clf;
plot(MR,MP,'-o');
hold on;
plot(HR,HP,'r+');
% Ploting human segmentations' mean
plot(mean(HR),mean(HP),'go','MarkerFaceColor',[0 1
0],'MarkerSize',7);
axis equal; axis([0 1 0 1]);
xlabel('Recall'); ylabel('Precision');

% find best MF and HF-measure (should interpolate)
[t,idx] = max(mf(:));
[ht,hidx] = max(hf(:));
title(sprintf('MF=%.2g at (R,P)=(%.2g,%.2g) t=%.2g HF=%.2g at
(R,P)=(%.2g,%.2g) t=%.2g',...
              mf(idx),MR(idx),MP(idx),mthresh(idx),
hf(hidx),HR(hidx),HP(hidx),hthresh(hidx)));
plot(MR(idx),MP(idx),'bs','MarkerFaceColor',[0 0 1],'MarkerSize',7);
```

```
plot(HR(hidx),HP(hidx),'rs','MarkerFaceColor',[1 0
0],'MarkerSize',7);
hold off;

end
```

machinePR.m

```
function [p,r,f,thresh] = machinePR(present, iid)

% This code is responsible for performing the tests required
% to gather results for further analysis - PR graphs

% setup
% present = 'color';
% iid = 800010;
nthresh = 15;

% read the image --- in BSDS/images
% im = rgb2gray(double(imread(imgFilename(iid)))/255);
im = double(imread(imgFilename(iid)))/255;
% figure(1); clf;
% imshow(im);

% create a pb image
pb = pbBGTG(im);

% Getting CG as well
pb1 = pbCGTG(im);

% Combining two detectors into one pb
[m,n] = size(pb);
for i=1:m
    for j=1:n
        if (pb(i,j)<pb1(i,j))
            pb(i,j) = pb1(i,j);
        end
    end
end

% Plotting the final results in pb
figure(2); clf;
imagesc(pb, [0 1]);
axis image; axis off; truesize;

% read segmentations --- in BSDS/human/present
segs = readSegs(present,iid);

% match the first seg and a thresholded pb
bmap1 = double(seg2bmap(segs{1}));
bmap2 = double(pb > 0.5);
[match1,match2,cost,oc] = correspondPixels(bmap1,bmap2,0.01,1000);
```

```
h=figure(3); clf;
plotMatch(h,bmap1,bmap2,match1,match2);
title('Seg1 vs. Pb>0.5','Color',[1 1 1]);

% compare the pb and segs
[thresh,cntR,sumR,cntP,sumP] = boundaryPR(pb,segs,nthresh);

% precision/recall plot
r = cntR./(sumR+(sumR==0));
p = cntP./(sumP+(sumP==0));
f = 2.*r.*p./(r+p+((r+p)==0));

end
```

humanPR.m

```
function [p,r,f,thresh] = humanPR(present, iid, uid)
% This function is responsible for calculating the PR
% values of a particular human segmentation compared to
% the rest.

% setup
nthresh = 1;

% read the seg file --- in BSDS/human/present
file = segFilename(present,uid,iid);
pb = seg2bmap(readSeg(file));

% Read segmentations --- in BSDS/human/present
segs = readSegs(present,iid);

% Remove the current seg file being tested from segs variable
segs(uid-1000) = [];

% compare the pb and segs
[thresh,cntR,sumR,cntP,sumP] = boundaryPR(pb,segs,nthresh);

% precision/recall calculation
r = cntR./(sumR+(sumR==0));
p = cntP./(sumP+(sumP==0));
f = 2.*r.*p./(r+p+((r+p)==0));

end
```